MANUAL OF MENTAL AND PHYSICAL TESTS

Part 1: Simpler Processes

THIRD EDITION INCLUDING SUPPLEMENTARY REFERENCES 1914 to 1919

MANUAL OF MENTAL AND PHYSICAL TESTS

Part I: Simpler Processes

A BOOK OF DIRECTIONS
COMPILED WITH SPECIAL REFERENCE TO THE EXPERIMENTAL STUDY
OF SCHOOL CHILDREN IN THE LABORATORY
OR CLASSROOM

BY

GUY MONTROSE WHIPPLE, PH.D.

UNIVERSITY OF MICHIGAN

AUTHOR OF "A GUIDE TO HIGH SCHOOL OBSERVATION," "QUESTIONS IN GENERAL AND EDUCATIONAL PSYCHOLOGY" "QUESTIONS IN SCHOOL HYGIENE"

BALTIMORE, U. S. A.
WARWICK & YORK, Inc.
1914

PREFACE TO THE SECOND EDITION.*

One need not be a close observer to perceive how markedly the interest in mental tests has developed during the past few years. Not very long ago attention to tests was largely restricted to a few laboratory psychologists; now tests have become objects of attention for many workers whose primary interest is in education, social service, medicine, industrial management and many other fields in which applied psychology promises valuable returns. It is a source of gratification to me that this book has been a positive factor in furthering this development of mental tests, and that it has found sufficient favor, both with academic investigators and teachers and with those engaged in the immediately practical task of mental examination, to exhaust the first edition in an unexpectedly short time.

Opportunity has been taken of this necessity for a new edition to make extensive revision and alterations in the text. The attempt has been made to eliminate a number of confusing typographical errors; a new and better system of abbreviations for titles of periodicals has been introduced; the directions for conducting the tests have been modified where ambiguity was felt; numerous tables and graphs have been added; the section upon the treatment of measures has been expanded; and several new tests have been inserted. These changes have so increased the material as to make it desirable to divide it into two parts, the first dealing with simpler, the second with more complex processes. The numbering of the original tests has, however, been retained for various reasons.

It would have been quite difficult for me to have undertaken to improve the text had I not received valuable assistance from many of my colleagues and co-workers. In especial I wish to

*For additional references from 1914 to 1919 see Appendix IV, References Since the Second Edition, page 353.

record my indebtedness to Dr. Otto Bobertag, of the *Institut für angewandte Psychologie,* Professor C. Spearman, of London, Professor S. P. Hayes of Mount Holyoke College, and Mr. D. K. Fraser, of Cornell University.

G. M. W.

Ithaca, N. Y.,
February, 1914.

PREFACE TO THE FIRST EDITION.

Hitherto the literature of mental and physical tests has been scattered in numerous journals; the results obtained by different investigators have too often not been compared; indeed, in many cases where the methods have been divergent, comparison has been impossible. In consequence, there have been no recognized standards of procedure and none of performance. Nevertheless, I believe that the time has now come for the taking of an account of stock, and for the systematization of the available materials. This conviction, which is the outgrowth of my own interest in the experimental study of mental capacities, an interest that has been with me during the past ten years, has been confirmed by many suggestions from colleagues and friends, who have pointed out that a manual of directions for mental tests would meet a real need, and might further the cause of investigation. More particularly, at the instigation of Mr. C. H. Stoelting, of Chicago, who has undertaken to supply the apparatus and materials prescribed in this volume, I began, in March, 1906, to prepare a small handbook of mental tests. The impossibility of adequate treatment of the subject in small compass has, however, necessitated the expansion of that early undertaking into the present work.

In the introductory sections of this volume, I have sought to show the general purposes of mental tests, to lay down rules for their conduct, and to explain the methods of treating data. In this connection I discuss the calculation of measures of general tendency, measures of variability, indexes of correlation, and other statistical constants.

In the body of the volume, I have brought together, for specific treatment, some fifty of the most promising tests. In every case, my plan has been to sketch the development of the test, to prescribe a standard form of apparatus and method of

procedure, to explain the treatment of the data secured, and to set forth the results and conclusions thus far obtained.

The tests that I have selected may not prove, ultimately, to be those of most value, but they are, I think, numerous enough, and varied enough in type, to furnish a working basis for investigations for some time to come.

In the choice of materials and methods, I have sought to follow a middle course; on the one hand to avoid the use of costly instruments of precision and of the elaborate methodology of the psychological laboratory, and on the other hand, to avoid the inexactness of make-shift apparatus and the unreliability of casual, unsystematic observation. My idea has been to supplement the exposition of the standard apparatus and method of procedure by suggestions for variations of apparatus or of method, so that each test will be carefully standardized, yet will retain a sufficient degree of flexibility.

Doubtless, to some readers, the instructions for the conduct of the tests will seem unnecessarily lengthy and detailed; but experience has convinced me that faulty results are to be traced, in quite the majority of instances, to the neglect of some seemingly trivial detail in the arrangement of the experimental conditions; so that instructions can scarcely be made too explicit in a manual of directions in which standardization is the object.

In explaining the treatment of data, my aim has been to make clear the arithmetic of the various formulas, without insisting, in every case, upon acquaintance with the mathematical reasoning upon which the formula is based.

And when I speak of "the results and conclusions thus far obtained," I speak with the intent to make clear what, I am sure, is made evident more than once in the text, that this book presents, not a closed chapter in the experimental investigation of mental activity, but a program of work to be done.

My acknowledgments for aid should be numerous and ungrudging. These have been made in part in the text, but in many instances, material assistance has, perforce, gone without explicit acknowledgment. I wish, however, to make clear my indebtedness to C. H. Stoelting Co., for the loan of numerous

cuts, to Dr. Guy L. Noyes, of the University of Missouri, for assistance in the tests of vision, to Dr. H. H. Goddard, of Vineland, N. J., for the adaptation of the Binet-Simon tests to American conditions, to my colleague, Professor I. M. Bentley, as well as to my wife and to my mother, for the reading of proof, and to my colleagues, Professors Charles DeGarmo and E. B. Titchener, for almost daily advice and encouragement.

The inscription of the book to Professor Titchener is in token of my special debt to him as a teacher and as an expositor of the scientific method of attack in the solution of the problems of mental life.

GUY MONTROSE WHIPPLE.

Cornell University,
June, 1910.

KEY TO SUPPLEMENTARY REFERENCES

Supplementary References, Appendix IV, page 352, is to be consulted in connection with the following pages:

p 4	106	p 223	105
73	5,8,32,39,74 94,108,113, 124,130,132	230	105
		235	105
		243	105
87	32,95	261	105
91	32,95	278	45
100	32,39,93,94 113,130,132	290	12,28,33,34 36,77,104
129	62	305	12,17,37,39 51,110,121
147	7,11,12,31,45 56,57,68,70 109,124,132	329	3,11,12,22 27,31,32,40 54,56,63,68 71,80,90,91 92,97,111,127 131,132
151	60,78		
155	31,55,56,57		
160	11,55,132		
175	89,130		
193	109	338	88,129
200	19,48	340	31,56,57,101
213	89,105	348	11

Additional abbreviations--

JApPs: The Journal of Applied Psychology (Worcester, Mass.)

JEPs: The Journal of Experimental Psychology (Princeton, N.J.)

ScSoc: School and Society (Garrison,N.Y.)

TABLE OF CONTENTS

INTRODUCTORY

THE TESTS

CHAPTER VII. TESTS OF ATTENTION AND PERCEPTION

INDEX OF FIGURES

INDEX OF TABLES

INTRODUCTORY

CHAPTER I

The Nature and Purpose of Mental Tests[1]

When we speak of a mental test we have in mind the experimental determination for a given individual of some phase of his mental capacity, the scientific measurement of some one of his mental traits.

The mental test in some respects resembles, in some respects differs from the typical research-experiment of the psychological laboratory. Like this latter, the test is superior to the casual observation of everyday life because it is purposeful and methodical: it thus possesses all the merits common to experimental investigation at large, viz: the control of conditions (including the elimination of disturbing, and the systematic isolation of contributory factors), the possibility of repetition, and the possibility of subjecting the obtained results to quantitative treatment.

The primary difference between the research-experiment and the test-experiment is really one of aim. The test has a diagnostic, rather than a theoretical aim: its purpose is not to discover new facts, principles or laws for the science of psychology —though such a result may indirectly be attained—but to analyze, measure and rank the status or the efficiency of traits and capacities in the individual under examination. Again, unlike the typical research-experiment, the mental test ordinarily places little or no emphasis upon introspective observation by the subject, in part because of its relatively short duration, in part because it is frequently applied to inexperienced

[1]The tests with which this volume is concerned are mainly mental tests. Since, however, the intimacy of the relation between mind and body makes it well-nigh imperative to study their interrelations, attention has been paid to the more important anthropometric measurements and to those tests of physical capacity that have most frequently been used in the search for correlations of psychical and physical ability.

subjects who are incapable of aught but the most elementary introspection, but more especially just because it is concerned less with the qualitative examination or structural analysis of mental processes than with the quantitative determination of mental efficiency; because, in other words, it studies mental performance rather than mental content.

There is, however, danger of laying too much stress upon this distinction between quantitative and qualitative examination. Those, like Myers and Andrews, who see in mental tests nothing but statistics of performance, and who contend that in the absence of introspection it is impossible to get any information upon mental content, have failed entirely to appreciate how significant for qualitative exploration are tests when handled by a competent investigator who puts less emphasis upon mere accumulation of figures than upon the patient search for qualitative aspects of mental life implicit in the outcome of his investigations. To disregard these possibilities of the test-experiment is not only a loss, but also a positive source of danger to the progress of science, for it tends to suggest, if not to encourage the superficial execution of the tests themselves, especially on the part of untrained laymen—witness the present tendency to hand over to classroom teachers or psychologically untrained medical inspectors the diagnosis of the mental status of school children by the Binet-Simon tests.

The purposes for which mental tests have been developed are, of course, varied, but, roughly speaking, we may distinguish a theoretical interest on the part of laboratory psychologists, and a practical interest on the part of those who are concerned with mind at work in everyday life.

Historically, it appears that most of the tests now in use have originated in the psychological laboratory, either in the natural course of the development of experimental psychology as a system, *e. g.*, the usual tests of sensory discrimination, or as a consequence of special attempts to study mental capacity, particularly the interrelations of various mental capacities and of mental with physical capacities. It is, we think, not too much to hope that in time the application of mental tests will bear rich fruit in this field. We may hope that the skillful study of

mental functions by the test-method may supply us with a satisfactory account of the nature and interrelations of mental functions, just as the typical introspective experiment has been able to furnish an account of the structural make-up of mind. If we could, to take an instance, obtain an exact science of mental functions so that we could know the unit-characters of mind as the biologist knows, or expects to know, the unit-characters of plants and animals, the study of mental inheritance would be carried appreciably forward.

Outside the laboratory an active and very natural interest in mental tests has been exhibited by those who are busy with practical problems to the solution of which the scientific study of mind may be expected to contribute. It is, naturally, the educator to whom the development of a significant and reliable system of mental tests would most appeal, since he is concerned with the development of just those capacities of mind that these tests propose to measure.

It was a practical educational problem, for instance, that motivated the work of Binet and Simon, whose system of serial graded tests is just now to the forefront of popular interest. Of late, too, hopeful beginnings have been made in the application of mental tests to vocational guidance, whether in the selection of people for positions or the selection of positions for people. But the magnitude and complexity of these tasks, particularly of the second, is far greater than commonly supposed.

And this leads us to point out that there has been, unfortunately it seems to us, a disposition in some quarters to speak as if a science of mental tests were already achieved; as if, for instance, a child's native ability could now be measured as easily as his height, or as if his suggestibility or his capacity for concentration of attention could be determined as readily as his skull circumference or his breathing capacity. To make such assertions is surely misleading, for, as the study of the tests herein embodied will show, there is, at the present time, scarcely a single mental test that can be applied unequivocally as a psychical measuring-rod. The fact is we have not agreed upon methods of procedure; we too often do not know what we

are measuring; and we too seldom realize the astounding complexity, variety and delicacy of form of our psychical nature.

Paradoxical as it may seem, these are the reasons, we believe, that render the elaboration of a scientific system of mental tests a possibility, for, if the all-too-evident lack of agreement in the results of the investigations already made is not attributable to faulty or divergent methods, or to clumsiness and ignorance—if, in other words, the discrepancies are inherent and ultimate—then we never can have a science of mental tests.

What we need is not new tests, though they are welcome enough, but an exhaustive investigation of a selected group of tests that have already been described or proposed. In particular, we need more than anything else, at least from the point of view of application, the establishment of norms of performance for these tests—norms that are based upon investigations in which standard and prescribed methods of procedure have been followed in a rigid and undeviating manner. We need, for every mental test that has proved its worth, data from a sufficiently large number of individuals of all stages and types of mental development to supply a standard table of distribution by percentile grades like the tables already available for some of the commoner anthropometric measurements.

This book is an attempt to assist in the realization of this need. It presents a program of work, rather than a final system of results.

REFERENCES.[2]

(1) B. R. Andrews, Auditory tests. *AmJPs*, 15:14-56.

(2) L. P. Ayres, Psychological tests in vocational guidance. *JEdPs*, 4:1913, 232-7. (Also published separately, Russell Sage Foundation, New York City.)

(3) H. Münsterberg, Psychology and industrial efficiency. Cambridge, 1913.

(4) C. S. Myers, The pitfalls of 'mental tests.' *BrMedJ*, Jan. 29, 1911. (Also published separately. Pp. 8.)

(5) W. Stern, Die differentielle Psychologie in ihren methodischen Grundlagen. Leipzig, 1911. Especially Ch. 6.

(6) W. Stern, Die psychologischen Methoden der Intelligenzprüfung und deren Anwendung an Schulkindern. Leipzig, 1912. (Eng. trans. to appear in *EdPsMon*, Baltimore.)

(7) F. L. Wells, The principle of mental tests. *Sci*, 38:1913, 221-224.

[2]Consult the *List of Abbreviations* for the exact titles of the periodicals cited in these and subsequent references.

CHAPTER II

General Rules for the Conduct of Tests

The following general rules may be laid down at the outset:

(1) The essential and fundamental principle underlying the conduct of scientific tests is the *standardization of conditions.* This does not mean that expensive apparatus or instruments of precision are always necessary, but simply that the conditions under which a test is given to one person or to one group of persons must be identically followed in giving the same test to another person or group. We cannot always make the conditions ideal, but we can at least try to keep them constant. *If the conditions are varied, they must be varied intentionally and for a definite purpose.*

(2) *No detail in the 'setting' of a test is too trivial to be neglected.* This is, of course, merely a restatement of the previous principle in another form. It is noteworthy that the lack of accordance between the results obtained by different investigators in the use of what is ostensibly the same test almost invariably turns out to be due to seemingly trivial variations in the method of administering the test.

In particular, attention may be called here to such matters as the time of day at which the experiment is made, the nature of the instructions that precede the test, the emotional attitude of the participants toward the investigation, their ability exactly to comprehend what is wanted of them (of which more hereafter) and their willingness to do their best throughout the test. It is well to write out the preliminary instructions and to memorize them, after first making a trial in order to see if they are perfectly intelligible. Thus, for instance, to say to one class of school children: "Cross out all the *e*'s on this paper while I take your time with a watch," and to another class: "Cross out all the *e*'s on this paper as fast as you can" may mean the same thing to the experimenter, but it will not bring the same results from the classes under investigation, because in the second case the idea of fast work has been more strongly emphasized.

A special difficulty that arises here may be too readily overlooked by the untrained examiner. The subject of a test is thrown by the examiner's instructions into what may be termed a 'set' or an 'attunement' for the work before him. Now, even if each subject be given precisely the same verbal directions, it is entirely possible that these directions may be interpreted differently by different subjects. Moreover, this attunement is quite as likely to be affected by the 'atmosphere' of the experiment, the general setting in which the work is carried on, as by the verbal instructions themselves. With children this difficulty may be quite serious. Anyone who has watched a succession of children enter the laboratory for mental testing must have been struck by these differences of attitude: one child is excessively timid, another is excited, a third is apathetic, a fourth is full of curiosity, and a fifth is strongly imbued with a competitive spirit. The skilled experimenter is on the watch for these signs and adapts his manner and his instructions in the effort so to compensate the tendencies that underlie them as to produce, so far as may be possible, the same final attunement.

The principle of rigid adherence to predetermined instructions has yet another exception. It may be quite worth while to vary instructions, like any other phase of the conditions of the test, provided the variation be done intelligently, for a purpose, by a trained experimenter. It is, indeed, precisely by this shifting of conditions that the first-class investigator can transcend the bare gathering of quantitative statistics and gain an insight into the qualitative aspects of mental life. Many boastfully 'exact' investigations would have more psychology in them had their authors known how to vary conditions and then, of course, to watch the results for concomitant variations in them. In short, then, the novice in mental testing neither realizes how disturbing slight changes in the experimental conditions can be, nor, on the other hand, how much of psychological interest can be discovered by intentional variation of the conditions.

(3) The examiner, *E*, should be on the look-out for external signs of the way in which the subject, *S*, responds to the test,

i. e., for indications of readiness, of quick comprehension, of a competitive spirit, or of *ennui,* fatigue, distraction, shift of attention, trickiness or deceit. The record-blanks should have a space for the recording of remarks of this nature. When tests are conducted individually, it is surprising how much can be gleaned in regard to *S*'s mental traits by these indirect hints. In particular, whenever the object of the test is to examine the correlation of some physical or mental trait with *S*'s general intelligence, it is largely upon this sort of observational record that *E* must depend for his estimate of this general intelligence, even though the test be supplemented by school marks, the estimates of teachers, and similar devices.

(4) No test should be undertaken by *E* until he is perfectly familiar with its nature, its purpose and its administration. Especially if it involves the use of apparatus, he should familiarize himself with the manipulations until they become automatic.

(5) No test should be undertaken until *S* is perfectly clear as to what is required of him. Since most mental tests are of an unfamiliar character, something beside explicit instructions, however clearly put, is needed to enable the average *S* to undertake the test under proper conditions. Ordinarily, a brief period (say 1 to 5 minutes) of *fore-exercise* is needed to remove timidity, excitement or misunderstanding. If this preliminary exercise is properly arranged (especially by being based upon material not used in the test proper, and by being of the same length and character for all *S*'s), it does not introduce a serious practise error, while it does decidedly facilitate the test. In some cases, however, as, for instance, when the facility of adaptation to the test-conditions is itself an object of investigation, the fore-exercise should be omitted.

(6) Most mental tests may be administered either to individuals or to groups. Both methods have advantages and disadvantages. The *group method* has, of course, the particular merit of economy of time; a class of 50 or 100 children may take a test in less than a fiftieth or a hundreth of the time

needed to administer the same test individually. Again, in certain comparative studies, *e. g.*, of the effects of a week's vacation upon the mental efficiency of school children, it becomes imperative that all *S*'s should take the tests at the same time.

But, on the whole, and especially when careful analytic work is contemplated, the group method, save for the preliminary trial of a method, is out of place. There are almost sure to be some *S*'s in every group that, for one reason or another, fail to follow instructions or to execute the test to the best of their ability. The individual method allows *E* to detect these cases. and in general, by the exercise of personal supervision, to gain, as has been noted above, valuable information concerning *S*'s attitude toward the test. Moreover, with the group method *E* must be content with bare quantitative performance: he has no opportunity for the skillful adaptation and variation of the attunement that we have mentioned; he can only surmise what has lain in *S*'s mind between instruction and performance, between stimulus and response, nor can he tell what effect the compulsion to work with other *S*'s may have had upon any given *S*. The objection that individual work takes too much time is, as one psychologist has put it, as laughable as would be the defence of a chemist that he had distilled several different fluids in the same flask without washing it—"to save time."

(7) One phase of the group *vs.* individual procedure demands special attention, viz: the problem of time-control. In many, if not in most tests, efficiency is measured, at least in part, by the rate at which the assigned work is performed. Now, in theory, rate or speed might be measured either by the amount of work performed within a given time or by the time taken to perform a given amount of work, in other words, by a *time-limit method* or by a *work-limit method.* In practise, however, we often find it difficult to arrange the material of a test in such a way as to make the task of equal objective difficulty at every portion of the test, and, even when this is possible, subjective variations may appear because of the fact that

different S's accomplish different amounts of work.[1] There is no doubt, therefore, that the work-limit method is to be preferred to the time-limit method: it is better, in other words, that every S should be asked to perform the same work and to measure his efficiency in terms of elapsed time than to require every S to work for the same time and to measure his efficiency in 'ground covered.' But the time-limit method is compulsory in all tests of this order undertaken by groups, so that this constitutes yet another serious objection to the group-method.[2]

(8) This leads naturally to a consideration of other difficulties that arise in scoring individual performance. Special difficulties are considered later in the discussion of the tests in which they appear, while the methods of handling measurements in general are treated in the following chapter. Attention is called here, however, to a fundamental problem, viz: *the relation of quantity of work to quality of work.* These two factors appear in nearly every test of mental efficiency, and the question arises: shall efficiency be measured in terms of quality, excellence, delicacy or accuracy of work, or shall it be measured in terms of quantity, rate, or speed of work? To this question no general answer can be given. Roughly speaking, quantity and quality of work probably tend, at least for a given S, to be inversely related. Whenever this relation can be demonstrated, it is theoretically, and often, indeed, actually possible to convert the two measures into a single index of 'net efficiency'—an index that is much desired for the study of general com-

---

[1]To take a concrete case, suppose E tries to conduct the "opposites" test upon a group of S's by the time-limit method. He must stop the work of all S's at the same moment. Suppose Subject A has written 10 and Subject B 20 opposites. Does it follow that B is twice as efficient as A? It certainly does not if, say, the 9th and the 10th terms on the list are more difficult than the rest, even if they are more difficult only to B. Again, it is more than twice as hard to name 120 as to name 60 words at random in three minutes. Failure to perceive this simple fallacy of counting as statistically equal, different portions or sections of an extended task is all too evident in the work of many investigators.

[2]Since the above was written the author has been able to use the work-limit method successfully with groups of reliable S's by the aid of a newly devised time-clock.

parative relations. In other instances it has been proposed[1] so to adjust the conditions of the test as to throw the emphasis so strongly upon quantity or upon quality of performance that the unemphasized factor may be neglected. In yet other instances, it seems necessary to keep both an index of quantity and an index of quality, and to make reference to both in subsequent comparative study.

(9) *Whether speed is a legitimate index of mental efficiency in general,* even when measured by the method of work-limits with individual *S*'s may perhaps be questioned. It is true that in many instances it seems most obvious, and is extremely tempting to compare the work of different *S*'s in terms of speed. But, in the author's opinion, we have been led astray by this temptation. Certainly, if we seek to evaluate the complex 'higher' mental functions, speed is not the primary index of efficiency, as is borne out by the evidence that speed and intelligence are not very highly correlated. Even with simpler tests, the quality of the work may be lost sight of, if rate of work is the only criterion of excellence: some *S*'s become feverishly and nervously excited and get the idea that the faster they work, the better; others are naturally critical and hesitate to proceed until they feel sure of themselves; others are habitually leisurely in their mental activities and will not 'speed up' unnecessarily. Again, the time recorded for performing a given task is often quite equivocal, because it includes a number of irrelevant factors whose quantum is unknown, particularly motor activities of various kinds. As a general rule, it may be said that *time measurements become more significant and reliable in proportion as the task becomes more mechanical and less intellectual.*

(10) As already intimated, the use of group-tests nearly always implies the use of *written responses.* When speed of performance is then used as a measure of efficiency, a perfectly obvious and often very serious source of error is introduced, for it is next to impossible, particularly with children, to measure or even to estimate the fraction of the total time that has been

---

[1] For an illustration of both of these methods for obtaining a single index, see the Cancellation Test.

expended upon the writing: the process of writing 'telescopes' to an unknown extent into the other activities set in play by the test. And, moreover, even when speed is not regarded as significant, most *S*'s will make a shorter and a different response to a test when they write than when they report orally.

For these reasons *mental tests should, so far as possible, be adapted for oral, instead of for written responses,* particularly if efficiency is to be measured in terms of speed. If written work be unavoidable, it should be reduced to the very minimum of simplicity.

(11) In the past not enough attention has been paid to the *desirability of repeating tests.* It is true that any repetition introduces the factor of practise and that this factor may exert a varying and undetermined influence. On the other hand, it is equally true that the outcome of a single application of a test may be very misleading: it may be affected by *S*'s special attunement, by his mood, by some chance accident of expression or by mere whim. Indeed, it has been recently argued by one writer that the low degree of correlation between various mental traits found by many workers in correlational psychology is directly due to their failure to repeat the tests until the initial irregularities have disappeared. And it would follow, in his opinion, that, while simple preliminary tests might serve very well for detecting extreme departures from the mode, "in the determination of individual differences within the large middle range of the curve of distribution * * * we shall find it necessary to determine the individual's 'limit of practise' in the various tests before we shall secure diagnostic results which will be verified by the individual's subsequent achievement in daily life."[1] Again, now that interest is directed so much toward the question of 'types,' it seems particularly necessary to caution against the fallacy of taking the

[1] H. L. Hollingworth, Correlation of abilities as affected by practice. *JEdPs*, 4:1913, 405-414. The reader will find a divergent conclusion expressed by C. Burt (Experimental tests of general intelligence, *BrJPs*, 3:1909, 94-177, especially 168-9), who declares "the correlation is highest on the first occasion, that is to say, when the task is newest"—referring to the correlation between experimental tests and general intelligence.

result of a single test as a positive indication that *S* falls into this or that type—because, of course, belonging to a type really implies the possession of a persistent tendency.[1]

(12) When it is desired to determine for a given *S* his efficiency in a given mental direction, *e. g.*, in memorizing, discriminating, observing, reasoning, synthetizing, inventing, and like functions, it must be borne in mind that, at least when the function is of a 'higher' and more complex order, *more than one test must be used.* The results of correlation psychology are proof that a series of tests that seem, all of them, to test the same mental function actually need not exhibit a high index of correlation. Nor can we regard any mental function as so clean cut, distinct and open to isolation that any single mental test can fully and finally map its form and dimensions. As Ebbinghaus has remarked: "There are no paragraphs in mind," and if there were, we might add, we could not record them by any single observation. *A fortiori,* to try to concoct a single and final test of such a comprehensive capacity as "general intelligence" becomes doubly absurd. Our general rule would be, then: use as many tests as possible and combine their data (quantitatively, if that prove feasible) to obtain a resultant value. Naturally, the tests must be selected with great care.

(13) In the application of any test, it is usual first to secure certain *preliminary data* concerning *S*'s personal history. Thus, in the experimental study of school children, *E* will find it advisable to record (*a*) name of the pupil in full, (*b*) sex, (*c*) date of birth, (*d*) name of school, (*e*) grade, (*f*) date, (*g*) hour. Other items, less uniformly recorded, but often of interest, are the following: (*h*) general health, (*i*) color of eyes and hair, (*j*) right or left-handedness, (*k*) name of teacher. (*l*) names and address of parents, (*m*) nationality of parents. (*n*) date of birth of parents, (*o*) occupation of parents, (*p*)

[1]An excellent illustration of the need of caution here is afforded by the work of Binet when he was studying the mental constitution of his daughter, Armande (*L'étude expérimentale de l'intelligence*). It will be recalled that Armande, on being asked why she now responded in a vein quite at variance with her previous replies, said: "I don't like to put things that way now; it would seem silly." That Binet, who was eager to establish the hypothesis of 'types,' had the courage to chronicle this incident, is one more testimony to the integrity of his scientific conscience.

number of children in family and their sex, (*q*) number of pupil in children of his family, (*r*) medical history of the pupil and his family, (*s*) obvious developmental defects or physical peculiarities, (*t*) details of personal habits, such as sleeping, eating, drinking, smoking, exercise, work, etc., (*u*) conduct in school, (*v*) proficiency in school work. For clinical work there will be added, of course, other anamnesic data appropriate to the case in hand.[1]

In *recording age* it is best to note the exact date of *S*'s birth. Unfortunately, direct comparison of the results of different investigators has at times been rendered difficult on account of disparity in the method of recording age. Thus, in arranging statistical tables, a boy 9 years and 7 months old would by some be classed in the group of 9-year-olds, by others in the group of 10-year-olds, as being nearer 10 than 9. A third method, which has the advantage of being clear to the reader and not confusing to *E*, is to put all *S*'s at or past a given birthday into a single group, the age of which is specified as that birthday, plus a half-year, *e. g.*, all *S*'s between their 9th and 10th birthday comprise the 9.5-year-old group, since their average age tends, of course, to approximate 9.5 years.

[1]For an excellent syllabus illustrating the data usually recorded for clinical examination of children, consult E. B. Huey, Backward and feeble-minded children. *EdPsMon*, Baltimore, 1912, ch. vi.

CHAPTER III

THE TREATMENT OF MEASURES

The immediate results of the application of mental and physical tests are very apt to be obscure or unintelligible until they have been ordered and systematized by proper statistical treatment. It is the purpose of the present chapter to explain the most common methods by which this systematization is accomplished.[1]

A. MEASURES OF GENERAL TENDENCY

In many cases it is unnecessary, if not impossible, to keep in view the individual measurements of an extended series. We naturally seek to condense these values into a single representative value. Any single measure that affords us such a summary of a series of measurements may be termed a 'representative measure' or a 'measure of general tendency.' There are three such measures in common use—the average or mean, the median, and the mode.

1. *The Mean*

(*a*) The *ordinary arithmetical mean* (M), more often termed the *average* in psychological measurements, is computed by dividing the sum of the several measurements or magnitudes (m) by their number (n).

Hence

$$M = \frac{\Sigma\, m}{n}. \tag{1}$$

[1]The reader will find more extended discussions of measurement methods in the following: Brown, Yule, Galton, Thorndike, Titchener, Sanford, Wissler, Spearman, C. B. Davenport, E. Davenport, Merriman and Elderton (see the end of this chapter for exact references). Technical papers upon correlation formulas by Pearson, Yule and others will be found in various numbers of *Biometrika*, the *Proc. of the Royal Soc.* of London, and in the *Phil. Transactions* of the same body.

The mean is the most familiar measure of general tendency, and it is the most precise, because it is affected by all measurements in proportion to their size. It has, however, some disadvantages: its computation requires more labor than that of the median or the mode, and, as will be shown later, it may fail after all to afford a truly representative value. Examples of arithmetical means are scarcely needed, but may be found in Tables 1 and 2, and elsewhere.[1]

(*b*) In dealing with a large number of measures a short cut may be furnished by a special application of what is known as the *weighted arithmetical mean.* Its use may be made evident by the following hypothetical case. Suppose it were desired to ascertain the average height of 1000 12-year-old boys. By the ordinary method we should be obliged to record each measure exactly (say, within 1 mm.) and to add the entire 1000 measurements. To utilize the weighted arithmetical mean, we divide the range of height into a limited number of groups of, let us say, 2 cm., and record simply the number of cases that fall into each group, *i. e.*, the frequency of each group. Thus in Table 3, there are in the 6th group 58 measurements lying between the limits 135 and 137 cm. The weighted mean can now be found very simply by multiplying the value or magnitude representing each group by the corresponding frequency (1×126; 5×128, etc.) and dividing the sum of the products by the sum of the frequencies (1000).

The formula for the weighted arithmetical mean is therefore

$$M = \frac{\Sigma\,(m \cdot f)}{\Sigma f},$$

or:

$$M = \frac{\Sigma\,(m \cdot f)}{n}. \qquad (2)$$

---

[1]The computation of M may be greatly lessened by assuming a convenient approximate value, and correcting subsequently to the true value. For illustrations, see Davenport (6, p. 429), Thorndike (32, ch. iv) or Yule (38, ch. vii).

TABLE 1

Strength of Grip, in Hectograms, 50 boys. (Whipple)

ORDER	RIGHT HAND				LEFT HAND				RANK COMPARISONS			
	NO.	STANDING	d	d^2	NO.	STANDING	d	d^2	G	D	D^2	xy
1	30	158	-125	15625	30	138	-135	18225	—	0	0	16875
2	17	175	-108	11664	17	163	-110	12100	—	0	0	11880
3	52	193	- 90	8100	1	175	- 98	9604	—	6	36	7470
4	39	197	- 86	7396	48	180	- 93	8649	—	2	4	7568
5	10	197	- 86	7396	7	180	- 93	8649	—	9	81	5246
6	43	200	- 83	6889	39	185	- 88	7744	—	2	4	6889
7	1	205	- 78	6084	16	190	- 83	6889	4	4	16	7644
8	40	206	- 77	5929	43	190	- 83	6889	—	4	16	5236
9	3	208	- 75	5625	52	190	- 83	6889	—	1	1	5550
10	7	210	- 73	5329	3	199	- 74	5476	5	5	25	6789
11	6	210	- 73	5329	6	200	- 73	5329	—	0	0	5329
12	48	220	- 63	3969	40	205	- 68	4624	8	8	64	5859
13	42	225	- 58	3364	15	210	- 63	3969	—	6	36	2204
14	2	225	- 58	3364	10	212	- 61	3721	—	2	4	2842
15	19	225	- 58	3364	19	215	- 58	3364	—	0	0	3364
16	37	226	- 57	3249	2	224	- 49	2401	—	10	100	741
17	15	235	- 48	2304	50	235	- 38	1444	4	4	16	3024
18	8	244	- 39	1521	8	235	- 38	1444	—	0	0	1482
19	14	244	- 39	1521	42	235	- 38	1444	—	6	36	507
20	51	245	- 38	1444	23	242	- 31	961	—	1	1	1102
21	50	248	- 35	1225	51	244	- 29	841	4	4	16	1330
22	41	262	- 21	441	45	245	- 28	784	—	2	4	273
23	25	262	- 21	441	9	253	- 20	400	—	14	196	- 735
24	23	267	- 16	256	41	260	- 13	169	4	4	16	496
25	12	269	- 14	196	14	260	- 13	169	—	5	25	- 98
26	44	270	- 13	169	37	260	- 13	169	—	1	1	78
27	29	273	- 10	100	44	267	- 6	36	—	2	4	- 20
28	9	280	- 3	9	34	270	- 3	9	5	5	25	60
29	45	290	+ 7	49	29	275	+ 2	4	7	7	49	- 196
30	36	294	11	121	12	280	7	49	—	10	100	594
31	11	296	13	169	32	282	9	81	—	3	9	312
32	28	301	18	324	20	290	17	289	—	4	16	576
33	32	310	27	729	13	290	17	289	2	2	4	243
34	31	313	30	900	11	297	24	576	—	1	1	810
35	20	315	32	1024	31	300	27	729	3	3	9	544
36	34	320	37	1369	28	305	32	1024	8	8	64	- 111
37	24	323	40	1600	25	308	35	1225	—	4	16	2520
38	16	325	42	1764	26	315	42	1764	31	31	961	-3486
39	35	330	47	2209	38	325	52	2704	—	3	9	3854
40	13	346	63	3969	36	327	54	2916	7	7	49	1071

It is clear that this weighted mean approaches the ordinary mean in accuracy in proportion as the number of classificatory groups is increased.

TABLE 1 (Continued)

ORDER	RIGHT HAND				LEFT HAND				RANK COMPARISONS			
	NO.	STANDING	d	d^2	NO.	STANDING	d	d^2	G	D	D^2	xy
41	38	348	65	4225	24	336	63	3969	2	2	4	3380
42	46	350	67	4489	35	355	82	6724	—	3	9	7035
43	33	353	70	4900	49	362	89	7921	—	1	1	7140
44	26	375	92	8464	33	375	102	10404	6	6	36	3864
45	21	375	92	8464	46	378	105	11025	—	2	4	12236
46	18	403	120	14400	18	400	127	16129	—	0	0	15240
47	27	430	147	21609	21	406	133	17689	—	2	4	26019
48	49	440	157	24649	5	443	170	28900	5	5	25	13973
49	5	440	157	24649	27	450	177	31329	1	1	1	26690
50	22	508	225	50625	22	490	217	47089	—	0	0	48825
Sums		14164	3104	293004		13651	3165	315221	106		2098	280118
Aver.		283	62.1	5860.1		273	63.3	6304.4				

TABLE 2

Values Derived from the Data of Table 1

MEASURE	FORMULA	VALUE	
		Right Hand	Left Hand
Mean	1	283.0	273.0
Median	3	269.5	260.0
A. D.	4	62.1	63.3
S. D.	5	76.6	79.4
S. D.	6	77.3	80.2
S. D.	7	77.8	79.3
P. E.	10	51.7	53.6
P. E.	11	52.1	54.2
P. E.	12	50.7	52.0
P. E.	13	52.5	53.5
C.	16	0.27	0.29

TABLE 3

Distribution of the Heights of 12-Year-Old Boys (Hypothetical)

Centimeters	126	128	130	132	134	136	138	140	142	144	146	148	150	152	154	156	158	160
Deviation	-16	-14	-12	-10	-8	-6	-4	-2	0	2	4	6	8	10	12	14	16	18
Frequency	1	5	14	24	39	58	96	120	150	142	123	88	63	36	23	12	5	1

Weighted mean = 142.9   Median = 142.9   Mode = 142.

2. *The Median*

The *median* or *central value* is, literally, the middlemost of a group of measurements arranged singly in ascending or descending order, or the measure above and below which lie an equal number of individual measurements. It is expressed, therefore, by the formula:

$$\text{median} = \text{the } \frac{n+1}{2} \text{ measurement.} \qquad (3)$$

In practise the median may or may not coincide with some actual measurement; more often than not it is an interpolated value. To compute its value we must first arrange the measures serially (if any magnitude is repeated two or more times, the number of such repetitions must, of course, be indicated). To find the middlemost measure when interpolation is needed, we may proceed by a simple method which may be illustrated by reference to Table 3. Here, since there are 1000 measurements, we seek the value of the 500.5th measurement. The first 8 groups (126 to 140 cm.) represent 357 measurements. The value desired, the 500.5th measurement, therefore lies in the 9th group of 150 cases and is the theoretical value between the 143d and the 144th measurement in this group. We have, therefore, to take $\frac{143.5}{150}$ of the range of magnitude covered by the 9th group (2 cm.), and add this to the lower limiting value of the group, 141 cm. (because the value 142 comprises measures ranging from 141 up to 143 cm.), so that we obtain for the median the value, $141 + 1.9 = 142.9$.

The great merit of the median is the ease with which it can

be determined: in short series it is not even necessary to arrange the measures serially, as one-half of the measurements may be checked off by inspection. Its primary disadvantage is that it gives little weight to extreme deviations and may fail entirely to represent the type, yet, in many psychological observations, it is precisely these extreme deviations which are most suspicious, so that this tendency of the median to lessen the significance of extreme measures may prove a positive advantage.

In general, the longer the series or the more symmetrical the distribution of its values, the more nearly does the median approximate the mean.

3. *The Mode*

If a number of measurements are distributed in ascending or descending order, a *mode* is a measure that appears more frequently than do measures just above or below it in the series. There may be several modes in a distribution, though usually there is but one, and we may then define the mode as the commonest single value, or the commonest condition.

Many statistical arrays find a better representative value in the mode than in the average. Thus, when we speak of the "average American citizen," we really have in mind the typical citizen, the one most frequently met with. To borrow an illustration from Rietz (6, p. 684): "If a community has 10 millionaires, but all the other citizens are in poverty, an arithmetical average might give the impression that the people of the community are in good financial condition, while really the 'average citizen' is in poverty." The primary use of the mode is, therefore, *to characterize a type*.

Strictly speaking, we may have an empirical mode, as indicated in a given array, and a theoretical mode, which would be the most frequent condition in a theoretical distribution. The latter is difficult to compute and not often employed. If an array is very irregular, there is, in strictness, no mode or type at all or at least the indicated mode has little significance.

In Table 3, it is clear that the mode is 142 cm., because this measure appears 150 times, and no other measure is so frequent.

B. MEASURES OF VARIABILITY

It is a common fallacy to rest content with the statement of the general tendency of measurements. Even in supposedly accurate and scientific determinations, we may find the quantitative expression limited to averages, *e. g.*, "the mean temperature for September," "the average weight of 12-year-old boys," etc. But it is evident that the average gives no indication of the distribution of the individual measures from which it is obtained, no indication of the extent to which these measures vary or deviate from the average, no information as to how homogeneous is the material that the average represents. The September temperature may have been seasonable and equable or there may have been some days of frost and some days of sweltering heat. Again, if five individuals weigh 80, 65, 60, 40, and 55 kg., respectively, and five others 62, 59, 60, 61, and 58 kg., respectively, then the mean weight of either group is 60 kg., but one group is distributed very closely around the mean, whereas the other group exhibits such marked deviations from it that *M* (or any other general tendency measure) has little or no significance as a representative value.

From this it follows that we need not only measures of general tendency, but also measures of the variability or tendency to deviation of measurements, and that these latter are of well-nigh equal importance.

There are three common measures of variability[1]—the average deviation, the standard deviation, and the probable error.[2]

[1]Besides these measures, range of variability is sometimes indicated roughly by stating the maximal and minimal measurements, in conjunction with *M*. This gives us, at least, information as to the extremes of deviation.

[2]It is well to avoid confusion here at the outset. The average deviation (A. D.), as used by the statisticians, is identical with the mean variation (m. v.) of experimental psychology. The standard deviation (σ) is called the average error by Sanford, the mean error by Merriman, and the error of mean square by others.

1. *The Average Deviation* (*Mean Variation*)

The average deviation, *A.D.*, of a series of measurements, m, is the arithmetical mean of their separate deviations, d, from their mean, M (or some other measure of central tendency), taken without regard to sign.

Hence:

$$A.D. \text{ or } m.v. = \frac{|M - m_1| + |M - m_2| + \ldots\ldots |M - m_n|}{n},$$

or:

$$A.D. = \frac{d_1 + d_2 + \ldots d_n}{n},$$

or:

$$A.D. = \frac{\Sigma\, d}{n}. \qquad (4)$$

Reference to Table 1 will render this process clear: there the average right-hand grip is 283; the weakest boy has a grip of 158, hence he deviates 125 units from the average; the strongest has a grip of 508, hence he deviates 225 units from the average. The first 28 boys rank below average, and their deviations, for a reason that will be clear later, are considered in the table as minus deviations. So far as the *A.D.* is concerned, however, all the deviations are added without regard to sign and their sum, 3104, is divided by the number of cases, 50, yielding a mean variation of 62.1 hectograms. If the median were selected as the representative value, the variability would, of course, be computed similarly with a new series of d's.

If the individual deviations are not to be made use of subsequently, as in correlation work, a very neat short-cut to the computation of the *A.D.* may be made in the following manner.[1]

---

[1]See Dunlap (7) for further suggestions concerning the use of the adding machine in this method.

Let M = the mean,
N = the total number of measures,
N_{-M} = the number of measures less than the mean,
N_{+M} = the number of measures greater than the mean,
Σ_{-M} = the sum of the measures less than the mean,
Σ_{+M} = the sum of the measures greater than the mean.
Then $A.D.$ may be found by either of the formulas

$$A.D. = (\Sigma_{+M} - M \cdot N_{+M}) \div .5N, \qquad (4a)$$

$$A.D. = (M \cdot N_{-M} - \Sigma_{-M}) \div .5N. \qquad (4b)$$

Thus, for the right-hand grip, Table 1, Formula 4a becomes $A.D. = (7785 - 283 \cdot 22) \div 25 = 62.3$.

2. *The Standard Deviation (Error of Mean Square)*

This measure of variability is preferred by many experimenters and is practically the only one employed by statisticians, as it is thought to be more accurate than the average deviation; but it is much more laborious to compute. It is the square root of the average of the squares of the individual deviations.

$$S.D., \text{ or } \sigma = \sqrt{\frac{d^2_1 + d^2_2 + d^2_3 + \ldots d^2_n}{n}},$$

or:

$$\sigma = \sqrt{\frac{\Sigma\,(d^2)}{n}} \qquad (5)$$

If n is small, the formula is often modified by writing $n-1$ in place of n:[1]
Hence:

$$\sigma = \sqrt{\frac{\Sigma\,(d^2)}{n-1}}. \qquad (6)$$

The application of Formula 5 is illustrated in Table 1, 5th and 9th columns, where the squares of the individual deviations are shown in detail. The sum of these squares for the right-

[1] For the reasons for this substitution, consult Merriman (p. 71). It is evident that the effect of the substitution becomes progressively less as n increases: as will be seen in Table 2, the difference between Formula 5 and Formula 6 is practically negligible when $n = 50$.

hand grip is 293,004. This is divided by 50, giving 5860.1, the square root of which is 76.6, the σ desired.

The *S. D.* of a given series is somewhat larger than its *A. D.* Theoretically, and practically if the distribution be 'normal' (in the sense to be explained) and the observations sufficiently numerous, the relation is constant at

$$\sigma = 1.2533 \ A. D. \tag{7}$$

Conversely,

$$A. D. = 0.7979 \, \sigma \tag{8}$$

As shown in Table 2, the *S. D.* computed by Formula 7 is closely similar to that computed by Formula 6.

3. *The Probable Error*

The probable error, *P. E.*, of a single measure is an amount of deviation both above and below *M.* (or median or mode) that will include one-half of the individual measures; that is, it is a value such that the number of deviations that exceed it (in either direction from *M*) is the same as the number of deviations that fall short of it.[1]

The *P. E.* is approximately two-thirds the *S. D.*, or more exactly:

$$P. E. = 0.6745 \, \sigma \tag{9}$$

By reference to Formula 5 this becomes:

$$P. E. = 0.6745 \sqrt{\frac{\Sigma \, (d^2)}{n}}, \tag{10}$$

[1]The term 'probable error' is often a source of confusion to those unfamiliar with its use in mathematics. For a descriptive term, we might call it the 'median deviation,' since it is that deviation that is found midway from the representative value in either direction. The magnitude in question is not, of course, the *most* probable error, neither is it, in a certain sense, a 'mistake.' It is rather a 'probable sampling error:' we are unable to measure every possible instance of the thing we are studying, but must content ourselves with a restricted number of samples, usually so taken as to be 'random samples.' The *P. E.* serves to indicate the reliability of these random samples, the degree to which they probably depart from the true universal values. That the *P. E.* implies a symmetrical distribution of the samples is indicated in its definition and will be made clearer in a subsequent section.

or, for a small number of cases (Formula 6):

$$P.E. = 0.6745 \sqrt{\frac{\Sigma (d^2)}{n-1}}. \qquad (11)$$

In practise we may find the $P. E.$ approximately, if the distribution be assumed to be normal (see under D, below), by counting off one-fourth of the cases from either end of a series of measurements, and halving the difference between the two values thus found.

$$P.E. = \frac{m_{\frac{3}{4}n} - m_{\frac{1}{4}n}}{2}. \qquad (12)$$

Thus in Table 1, these limits lie at the 12th and a half and the 37th and a half measurements, and have, for the right-hand grip the values 222.5 and 324, respectively; hence, $P. E. = (324 - 222.5) \div 2 = 50.7$—a value that is approximately the same as the values of $P. E.$ computed by Formulas 10, 11, and 13 (Table 2) By Formula 11, $P. E. = 0.6745 \times 77.3 = 52.1$. Corresponding values are given in Table 2 for the left-hand grip as distributed in Table 1. Still other values might be computed on the basis of the median instead of the mean.

By combination of Formulas 7, 8, and 9, we may obtain for a normal distribution:

$$P.E. = 0.8453\, A.D. \qquad (13)$$

$$S.D. = 1.4825\, P.E. \qquad (14)$$

$$A.D. = 1.1843\, P.E. \qquad (15)$$

The first of these is illustrated in Table 2.

4. *The Coefficient of Variability*

If it is desired to compare the variability of one series of measurements with that of another, it will be found that, as a rule, their respective measures of variability cannot be compared directly, because they are based upon different units or at least upon different measures of general tendency, but the relations of the two measures of variability to their respective

measures of general tendency can be directly compared. In other words, we can compute two coefficients of variability (C) by dividing in each series a measure of variability by a representative measure, *i. e.*, either *S. D.*, *A. D.*, or *P. E.*, may be divided by either mean, median, or mode. Unless otherwise specified, it may be assumed that *S. D.* is divided by *M*. Hence:

$$C = \frac{\sigma}{M} \tag{16}$$

Thus, in Table 1, for strength of right hand, $C = 77.3 \div 283 = .27$ and for strength of left hand, $C = 80.2 \div 273 = .29$, hence the latter series is slightly more variable.

5. *Variability of Measures of General Tendency*

The measures of variability just discussed serve to inform us of the variability of the distribution of the individual measurements around the mean, mode or median. But since our opportunities for securing data are limited, it follows that even averages may fail to be absolutely exact measures of the general tendency of the trait under measurement. To revert to the hypothetical data of Table 3, we were there able to obtain an *M*, 142.9 cm., of the height of 12-year-old boys: it must be evident that if we could have measured a million boys we should feel surer that the *M* then obtained was the true one, or that if we had measured only ten boys of that age we should not have felt at all sure that the average thus obtained was truly representative of the height of 12-year-old boys. We need therefore, a measure of the reliability of *M*, so that we may have some idea as to how far the actually obtained *M* is likely to differ from the ideal or true *M*, or, reversely, how many measurements we need to secure an *M* that will have any desired or assigned degree of reliability.

To illustrate, suppose we did measure 1,000,000 boys in 1000 groups of 1000 measurements each; if we then averaged each group we should obtain 1000 *M*'s, each representing the central tendency of a group chosen by random sampling: we should

then expect these 1000 M's to be closely similar, but not identical, and we could distribute them like a series of individual measures and determine the variability of this distribution. In practise, the variability of M is found by a formula that takes into consideration the variability of the distribution which M represents and the number of cases on which it is based.

Any formula for the variability of a distribution can be converted into a formula for the variability of the mean of the distribution by dividing the formula in question by the square root of the number of measurements. Thus, if Formula 4 be expressed as $A.\,D._{\text{dis.}}$ and the average deviation of the mean by $A.\,D._M$, we obtain

$$A.\,D._M = \frac{A.\,D._{\text{dis.}}}{\sqrt{n}}. \tag{17}$$

Similarly, Formulas 5 or 6 yield

$$\sigma_M = \frac{\sigma_{\text{dis.}}}{\sqrt{n}}. \tag{18}$$

More frequently used, however, is

$$P.\,E._M = \frac{0.6745\,\sigma}{\sqrt{n}}. \tag{19}$$

6. *Variability of Differences between Means*

One of the commonest problems in the application of tests is the determination of a difference between the averages of two sets of data. Thus, the average capacity of boys might be compared with the average capacity of girls in a given test. The circumstance that the one average is higher than the other does not of necessity indicate a true difference: it is imperative first to determine the degree of reliability that attaches to the difference. The neglect of this obvious check is one of the most conspicuous faults in many published results. Yet the computation is perfectly simple: the variability of a difference between two means is the square root of the sum of the squares of the variabilities of these means. Any one of the measures of variability may be employed.

Let D be the difference between two means, M_1 and M_2: let n_1 be the number of cases in the first, n_2 in the second series of measurements and the remaining values be similarly designated by the subscripts 1 and 2.

By reference to Formula 17 we obtain

$$A.D._D = \sqrt{\left(\frac{A.D._{dis.1}}{\sqrt{n_1}}\right)^2 + \left(\frac{A.D._{dis.2}}{\sqrt{n_2}}\right)^2} \text{ or } \sqrt{\frac{A.D.^2_1}{n_1} + \frac{A.D.^2_2}{n_2}} \tag{20}$$

From Formula 18, similarly,

$$\sigma_D = \sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}, \tag{21}$$

and from Formula 19

$$P.E._D = \sqrt{\frac{P.E.^2_1}{n_1} + \frac{P.E.^2_2}{n_2}} \tag{22}$$

In case, however, the difference in question is a difference between the averages of two series of measurements that might stand in correlation (whether positive or negative, complete or incomplete), Formula 22 should be replaced by the following formula, in which r is the coefficient of correlation (explained later on) between the two series:

$$P.E._D = 0.6745\sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2}{n}}. \tag{22a}$$

In illustration[1] we apply this formula to the data of Table 1 in order to see whether the difference between right-hand and left-hand grip is a chance or a significant difference. Substituting for r its value 0.92 (as computed below) and for σ_1 and σ_2 the values found by Formula 6 (Table 2), we obtain P. E. = 3.0, or not quite one-third the magnitude of the difference, 10, which thus gains a satisfactory degree of reliability.

[1]For a more detailed but simple example of the application of this formula, consult W. H. Winch, Inductive versus deductive methods of teaching. *EdPsMon*, No. 11, Baltimore, 1913 (statistical note, pp. 7-10).

7. *Other Measures of Variability*

As in the case of the measures of general tendency and of differences in general tendency, so with the measures of variability, they likewise diverge from the true values because we are forced to limit ourselves to restricted or random samples of the facts we are recording. The degree of this divergence, *i. e.*, the reliability of measures of variability may be discovered, approximately at least, by dividing any one of these measures by $\sqrt{2n}$. To take a single case:

$$\sigma_{var.} = \frac{\sigma_{dis.}}{\sqrt{2n}} \tag{23}$$

C. THE GRAPHIC REPRESENTATION OF MEASUREMENTS

A series of measurements, as we have seen, can be expressed adequately by a single representative value only when that value is accompanied by some measure of variability. Even these two values may fail to express the series completely, since they are, after all, only symbols for the convenient summarizing of general tendency and variability, whereas a complete numerical expression of a series of measures would imply the tabulation of all the data of the series. Such a tabulation is for the most part impracticable, or at least of little significance, because of the difficulty of grasping the nature of the series by the inspection of a mass of figures.

The use of the graphic method, however, supplies a most serviceable and effective means of showing at a glance all of the important features in the distribution of a series of measurements and likewise of relations between series of measurements.

1. *The Plotting of Frequencies: Graphs of Distribution*

The most usual form of graph for illustrating the distribution of a series of measurements is constructed as follows:

Draw two lines, *OY* and *OX* (Fig. 1) in the form of coördinate axes, *i. e.*, with *OY* perpendicular to *OX*. Upon the hori-

zontal, or x-axis, lay off convenient intervals corresponding to the units of measurement of the series to be plotted; upon the vertical, or y-axis, lay off intervals corresponding to the frequencies of the series.

The choice of the scale units is largely arbitrary. The intervals of the two axes need not be the same, nor need different graphs, save for purposes of direct comparison, be plotted to the same scale. In general, a scale should be selected that will bring the surface easily into view as a whole and that will render conspicuous the features that are under consideration. Thus, if one is studying rate of increase or decrease, a scale should be selected that affords a fairly steep curve in order to emphasize its rise and fall. 'Squared' or cross-section paper (usually laid off by mm. on sheets 15x20 cm.) may be purchased for curve-plotting, and will be found invaluable for this work.

In illustration, the numerical table of frequencies above (Table 3) is turned into a surface of frequency upon the axes just mentioned (Fig. 1). We mark off on the x-axis, it will be seen, 18 equal intervals corresponding to the range of dimensions, 126, 128, ... 160 cm. Upon the y-axis we mark off equidistant intervals for the range of frequencies from 1 to 150. We next locate the series of 18 points. The first point lies vertically above the 126-cm. mark at a distance equal to 1 of the vertical units; the second lies vertically above the 128-cm. mark at a distance equal to 5 vertical units, etc. By joining the 18 points thus located, the resulting line evidently gives in a single visual impression the distribution that was expressed numerically in Table 3. Any point in this line is fixed by stating its abscissa or distance from the y-axis, and its ordinate, or distance from the x-axis.[1]

Now, it would have been equally feasible to have considered the values in Table 3 in terms of their deviation from the mean, median or mode, and with little or no change in the curve. Take, for simplicity, the mode, 142 cm., as the representative value. Erect an ordinate of the value of 150 at a point **M** on the x-axis (Fig. 1); intervals to the right of this ordinate may now represent positive deviations (+ 2, + 4 + 6, etc.) while those to the

[1] The 'curve' is sometimes so drawn as to form the tops of a series of columns erected at the intervals on the base-line, instead of by joining the single points as here described. See, for example, Fig. 2.

left represent negative deviations (— 2, — 4, — 6, etc.), as indicated in Table 3. It thus becomes possible to represent negative values graphically.

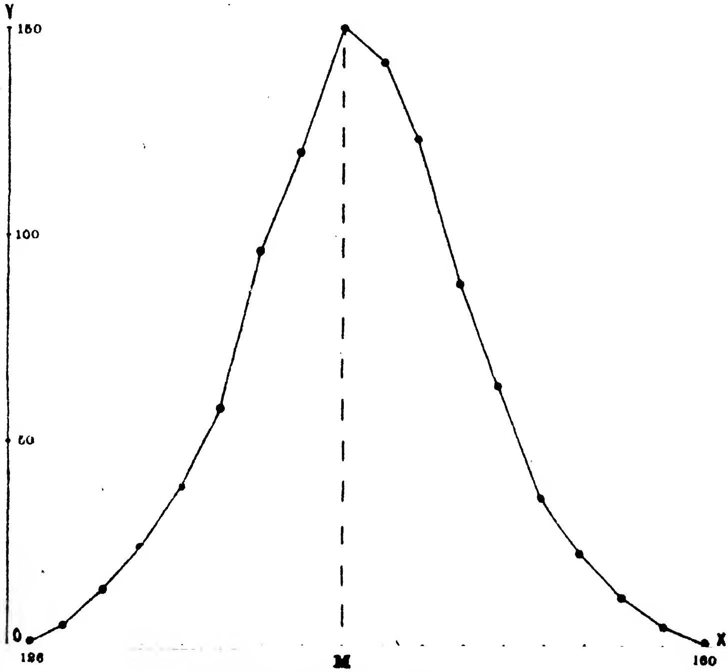


FIG. 1. GRAPHIC REPRESENTATION OF THE DISTRIBUTION OF TABLE 3.

2. *The 'Smoothing' of Distributions*

Ordinary measurements are subject to numerous disturbing factors; our units of measurement are often coarse; our opportunities for securing data are always restricted; variable factors of one sort or another obtrude themselves—and these disturbances produce irregularities in the resultant data. The obtained distribution, in other words, does not coincide with the true distribution, *i. e.,* with the distribution that would theoretically appear under ideal conditions. Thus, in Table 3. chance may have led to the measurement among the 1,000 cases studied, of more boys of a certain height, say 144 cm., than we should ordinarily have encountered in measuring 1,000 pupils

taken at random. Or, to take an instance of a striking artificial distortion, in the census returns, people who are 39 or 41 years of age show a tendency to report their age as 40, so that the age of 40 has an unnaturally large frequency.

Minor deviations from the theoretically expected distribution may be counteracted if we are constructing a frequency graph by 'smoothing' the curve, *i. e.*, by drawing the connecting line in the form of a true curve rather than a broken straight line: such a curve will pass in the neighborhood of the several points which have been located by the numerical data, but will not necessarily pass exactly through these points.[1] The result is a graph that shows how the data would presumably have been distributed if the factors which produced the distortions and irregularities were eliminated.

TABLE 4

The Numerical Smoothing of the Distribution of Table 3

Centimeters	126	128	130	132	134	136	138	140	142	144	146	148	150	152	154	156	158	160
Original	1	5	14	24	39	58	96	120	150	142	123	88	63	36	23	12	5	1
Smoothed	2	7	14	26	40	64	91	122	137	138	108	91	62	41	24	13	6	2

In tabular work these deviations may be counteracted by a simple arithmetical process. Replace each frequency except the two extreme ones by the mean (to the nearest integer) of the given frequency and the one just below it and the one just above it; replace the two extreme frequencies by the mean (to the nearest integer) of the given frequency taken two times and the adjacent frequency taken once. If necessary, a second smoothing may be made of the values obtained by the first smoothing.

The values of Table 3 do not exhibit marked irregularities, as is evident from their graphic distribution in Fig. 1: the process of smoothing may, however, be illustrated by the treatment in Table 4.

[1]For graphic work the curve may most readily be drawn with the aid of the celluloid 'curves' or the flexible splines sold for this purpose by dealers in drafting instruments.

D. NORMAL AND OTHER TYPES OF DISTRIBUTION: THE PROBABILITY SURFACE AND ITS APPLICATIONS[1]

1. *The Normal Frequency Surface*

Assume that errors of observation have been eliminated and that a large number of measurements of some psychological trait or capacity have been secured: experience has shown, and theoretical considerations likewise indicate, that as a rule these measurements will distribute themselves in the form of a symmetrical bell-shaped curve—variously known as the probability curve, the curve of error, Gauss' curve, or the normal frequency surface—the salient characteristics of which are a maximal frequency at M with a series of positive and negative d's from M that are symmetrically disposed on either side of it and whose frequency decreases progressively as their size increases.

Such a distribution implies the operation in the conditions that underlie each measurement of the feature or trait under measurement, of an indefinitely large number of individual factors, each of which is as likely to be present and effective in any one measurement as in any other. When, however, there are limiting or restricting conditions, or when one or more factors are present oftener than mere chance would allow, the resultant distribution will tend to depart from the normal type.[2] Thus the chances of death at different ages are not distributed according to the normal curve, but are higher in infancy and old age than in youth and middle age. The mental ability of college students is not likely to be distributed like that of the non-college population of the same age on account of the selective influence of entrance requirement. Thus, the curve of distribution for actual marks given to more than 20,000 Cornell University students (Fig. 2) is skewed to the right,

[1] For a detailed discussion of these topics, consult especially Yule, Ch. xv.

[2] It is, perhaps, hardly necessary to warn the reader of the distinction between (*a*) the number of observations and (*b*) the number of chance factors affecting each observation. If the former be large, the effect is merely to 'smooth' the distribution, but not necessarily to produce a normal distribution: if the latter be large, the distribution tends to approximate the 'normal' probability surface.

perhaps partly on account of this selective factor, but also on account of the operation of other disturbing factors, like the tendency to 'mark high,' the effect of zeal and industry in transmuting ability into the accomplishment that is being marked by the instructors, etc.[1] In general, distributions that do not conform to the normal type are termed 'skewed' distributions, and may demand special mathematical treatment.

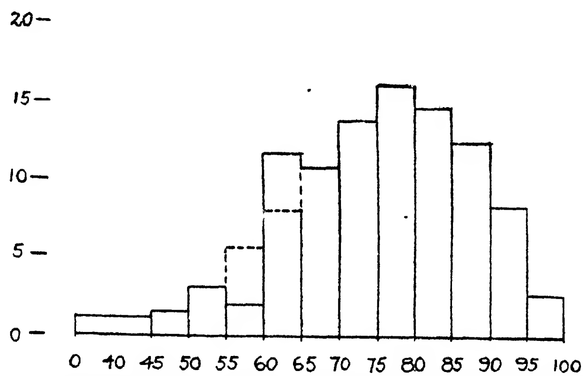


FIG. 2. DISTRIBUTION OF THE MARKS OF 20,348 CORNELL UNIVERSITY STUDENTS. From I. E. FINKELSTEIN. *The Marking System in Theory and Practice.*

2. *Relation of the Normal Curve to S. D. and P. E.*

The normal surface of frequency has interest still further because in it the significance of *P. E.* and of *S. D.* becomes clear. In fact, the latter bears to the curve a relation like that of a radius to its circle. If *S. D.* is small, the measurements are relatively similar and the curve is steep and compact (right-hand curve in Fig. 3), whereas, if *S. D.* is large, the curve is broad and of easy slope (left-hand curve in Fig. 3). If *M* and

[1]One special source of disturbance breaks the smoothness of the distribution at the 60 point, because this is the boundary between failure ("conditioned") and passing. Instructors dislike to report grades from 55 to 59, inclusive, but tend to change these to 60. This affords an excellent illustration of the manner in which a constant tendency will affect a curve of distribution. The dotted lines at this point in the diagram show what the curve theoretically should be were this constant error removed.

S. D. are known, the entire curve for a normal distribution is known. If the distribution is not of the normal form, the *S. D.* still remains a good measure of its variability, though not completely descriptive of the entire distribution.[1]

The geometrical explanation of the *P. E.* is simple. In Fig. 3 we draw the ordinates *ab* and *cd* equidistant from *OY* and at such a distance that the area *abYcd* is equal to the remainder of the total area under the curve: then the abscissa *Oa* or *Od* represents the value of *P. E.*, *i. e.*, a deviation above and below the mean that will include one-half of the total deviations.

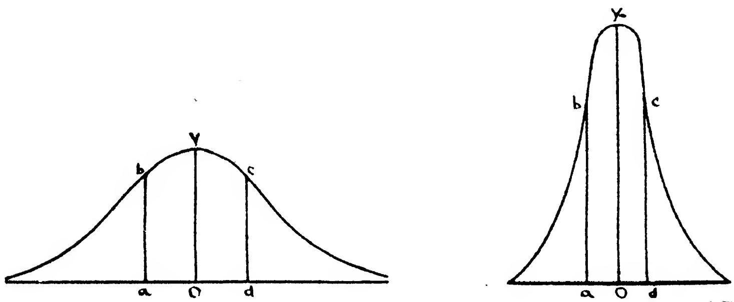


FIG. 3. TYPICAL CURVES OF NORMAL DISTRIBUTION.

The consistency of a series of measurements may also be indicated by stating the degree of probability that will attach to the appearance of an 'error' or deviation or residual, as it is often termed, of a magnitude equal to any assigned multiple of *P. E.* By definition, a deviation of the magnitude of *P. E.* is one as likely to be exceeded as not; in other words, the chances are even, or one to one, that it is exceeded. The probability of the occurrence of a deviation several times as large as *P. E.* is, however, very much smaller, as will be seen in the following

[1]This mathematical relation of *S. D.* to the probability curve, together with the possibility, as is shown later, of determining many other features of the distribution from the relation of *S. D.* and *M.* is one of the principal reasons why *S. D.* is preferred by many to the more-easily calculated *A. D.*

comparisons between P, the theoretical probability and $x \div P.E.$, multiples of $P.E.$, from 1 to 5.[1]

$x \div P.E.$	$P.$
1	$1 \div 1.0$
2	$1 \div 5.6$
3	$1 \div 23.2$
4	$1 \div 143.3$
5	$1 \div 1342.2$

E. MEASURES OF CORRELATION

1. *The Meaning of Correlation*

Physical science discovers numerous uniformities or correspondences between natural phenomena which are formulated as 'natural laws:' biological science, on account of the intricacy of the factors which condition vital phenomena, can discover, for the most part, only *tendencies* toward uniformity or tendencies toward correspondence. Such a tendency of two or more traits or capacities to vary together is termed a correlation. Thus height and weight are obviously correlated because in general tall people are heavier than short people, but, of course, this tendency toward correspondence is far from absolute like the correspondence between the distance and speed of a body falling in vacuum or between the electrical constants, voltage, amperage and resistance, as expressed in Ohm's law.

[1]These values are computed by reference to standard tables of values of the probability integral corresponding to various multiples of *P. E.* A condensed table of this sort is published by Thorndike (32, p. 200). The values given above were derived for the author by Prof. G. C. Comstock of the University of Wisconsin from Oppolzer's 10-figure table of the Gamma Integral, and are correct to the first place of decimals given. To illustrate from Thorndike's condensed table; the total area of the probability surface being 1000, the total area representing deviation in either direction is 500. From the table we see that a deviation or residual equal to 3 P. E. occurs in such a manner that 479 of the 500 cases are included between it and the average or median, and hence it is exceeded by 21 of 500 cases, or by 1 case in 23.8, approximately; 23.2 when more accurate integral tables are used.

From such a series of values the consistency of the determination may be stated in various ways. For example, if a correlation of .50 were accompanied by a *P. E.* of .10, it might be said that the chances would be but 1 in more than 1300 times that such a correlation would occur by mere chance.

Since in practically every psychological test we are searching for these tendencies toward correspondence, it is important to know how they can be measured. In not a few psychological investigations correlation has been expressed merely descriptively as 'fair,' 'large,' 'poor,' etc., and these characterizations have been derived from mere inspection of arrays of data. As a matter of fact, some of these published statements of correlation are actually wrong: correlations do not exist where they have been affirmed, or do exist where they have been denied. At the present time there is no excuse for such merely descriptive statements of correlation, since, by the use of appropriate mathematical procedure, a tendency toward correspondence may be measured and expressed by a single quantitative symbol that has as much significance and definiteness as *M, S. D.,* or any other statistical symbol. This symbol, *r,* which sums up the proportionality or degree of relationship between two factors or events, is known as the *index or coefficient of correlation.*

Complete positive or direct correlation between two traits is present when the existence of the one is invariably accompanied by the existence of the other, or when increase of the one is invariably accompanied by corresponding and proportional increase of the other.

Complete negative or inverse correlation is present when two traits are mutually exclusive, or when increase in the one is invariably accompanied by a corresponding and proportional decrease in the other.

A correlation is *indifferent* or *zero* if the existence or variation of one trait is totally unrelated to that of the other.

In perfect positive correlation, r is unity or 1.00; in complete negative correlation, r is -1.00; indifference or complete absence of correlation is 0. In actual psychological investigation, at least when functional correspondences are under investigation, we have commonly to deal with some intermediate degree of correlation, and r assumes, therefore, the form of a decimal lying between 0 and 1.00 for positive and between 0 and -1.00 for negative correspondence.

While the term 'correlation' is often broadly used in a generic sense to include any sort of relationship between two series of values, it is better to distinguish between it and other terms like 'dependency,' 'contingency' and 'association.'

The term 'dependency' properly applies to any functional relationship between two series, whether or not it can be expressed by the ordinary index of correlation. Such relationships may exist in so simple a form as to be readily expressed either verbally or mathematically, but whose coefficient of correlation would figure zero by the regular methods of computation. Take, for instance, the hypothetical relation that Spearman has adduced as an example of zero correlation (22, p. 77). Suppose five persons are tested for vision and hearing with the following results (in terms of feet that the test-type is read and the sound heard):

Person	*A*	*B*	*C*	*D*	*E*
Vision, in feet	6	7	9	11	14
Hearing, in feet	6	11	12	10	8

Here, $r = o$, and thus there is no correlation, direct or inverse. If, however, the values are plotted graphically, there is revealed a simple dependency, viz.: hearing is good when sight is poor, reaches a maximum when sight is fairly good, and then declines when sight continues to improve. (Lehmann and Pedersen, 16, p. 16.) From this it may be seen that it is best to plot relations in graphic form whenever possible. In many cases, indeed, such a functional graph is more significant than any coefficient could possibly be, just as a curve of distribution is more significant than an *M*, even when coupled with its measure of deviation.

If, however, by charting or otherwise, the relationship between the two arrays of data is found to be direct, if, in other words, the values of the one array vary proportionately with those of the other, there is not only dependency, but also correlation between the two arrays.

If these arrays each form a quantitatively graded series, the relationship may then be termed a '*series-correlation*.' If the gradations are expressed in numerical values, measures or magnitudes, the relationship is one of correlation in the specific sense of that term (*Masskorrelation* of the Germans). But if the gradations are expressed only in terms of rank or order in a series, the relationship is one of *rank-correlation* or *co-ordination* (to use the term proposed by Betz).

If the values do not form a quantitatively graded series, but are merely grouped into qualitative divisions, the relation is known as *contingency* (or, following Yule, as *association*, if the classification is limited to a two-fold division).

Examples of correlations, co-ordinations and contingencies appear in what follows. As already pointed out, these terms are not always distinguished, and the phrase 'degree of correlation' is often used to cover not only correlation proper, but also degree of co-ordination, of contingency and of association.[1]

[1] For further discussion of terminology and for the derivation of the formulas that follow, consult especially Betz, Stern and Yule.

2. *The Computation of the Index of Correlation*[1]

(a) The 'Product-Moments' Method of Pearson

The most elaborate as well as the theoretically best possible method of computing r is the standard 'product-moments' method elaborated by Bravais, Galton, and especially by Pearson.[2]

By the product-moments method

$$r = \frac{\Sigma\, xy}{n\, \sigma_1\, \sigma_2}, \tag{24}$$

in which the x's are the series of deviations from M in the first array and the y's the corresponding series of deviations in the second array, and in which σ_1 is the standard deviation of the first and σ_2 the standard deviation of the second array, and n the number of cases in either array.

The various steps of the computation may be illustrated by reference to Table 1[3] for grip of right and left hands, as follows:

(1) Arrange the original measurements in order of their standing or rank, as shown in Columns 1, 3, and 7. (While this is not absolutely necessary, it commonly facilitates computation, though for speedier determination of the xy values, it might be preferable to place the two arrays in the same order by individuals, *e. g.*, as shown by the numbers in Columns 2 and 6.)

(2) Compute M (or the median) for each series (283 and 273.)

(3) Compute and record the individual deviations (*d*, Columns 4 and 8) for each series, retaining the algebraic signs.[4]

---

[1] Consult Thorndike (32, chs. x and xi) for illustrations of other methods than those described here.

[2] In referring to the product-moments as the best possible method, certain qualifications must be kept in mind. It is possible, for instance, that, as Spearman contends (25), the comparison of ranks (*R*-method), for reasons that will be explained later, may be more reliable and satisfactory when psychological data are under treatment.

[3] For a fuller illustration of correlation arithmetic, together with suggestions for shortening the work, see E. Davenport, pp. 455-471.

[4] In Table 1, to accord better with common thinking, standings below average are considered as minus deviations, *i. e.*, the mean has been subtracted from the individual standings instead of vice versa.

(4) Multiply the d of each individual in the first series (now termed his x) by the d for the same individual in the second series (now termed his y), and record the products, observing the algebraic signs (the xy values in Column 13), *e. g.*, boy No. 30 has for his x, -125 and for his y, -135, hence, for his xy, $-125 \times -135 = 16875$. Again, boy No. 25 has for his corresponding values, -21 and $+35$, hence for xy, -735.

(5) Add the products obtained in (4) (280,118, Column 13).

(6) Compute the *S. D.* of both series (σ_1 and σ_2 in Formula 24, illustrated in the d^2 columns, 5th and 9th, and explained in Formula 5): multiply them together, and multiply their product by the number of cases ($76.6 \times 79.4 \times 50 = 304{,}102$).

(7) Divide the 5th by the 6th resultant for the index desired ($r = 280{,}118 \div 304{,}102 = +0.92$).

The arithmetic of the Pearson method is thus simple, though somewhat tedious. The work may be materially lessened by the use of Barlow's *Tables of Squares, Cubes, Square Roots*, etc., New York, 1904, of Crelle's *Rechentafeln* (procurable through G. E. Stechert & Co., New York), which show at a glance the products of all numbers up to 1000×1000, and by the use of an adding machine.[1] Another considerable shortening may often be effected without serious disturbance by substituting Formula 7 for Formula 5 in computing the two *S. D.*'s. Thus, in our illustration, this substitution (see Table 2) gives for the denominator of the fraction: $77.8 \times 79.3 \times 50 = 308{,}477$, from which we find $r = 0.94$.

The *probable error of the coefficient of correlation* as obtained by the Pearson method is calculated by the formula

$$P.E._r = 0.6745 \frac{1 - r^2}{\sqrt{n}}. \tag{25}$$

It is evident that the reliability of a coefficient increases with the number of cases compared and also with the magnitude of the r obtained. The actual values of $P.E._r$, as computed by Formula 25 for eleven values of r from 0 to 1, accompanying values of n from 25 to 1000, are indicated in Table 5, so that one can not only read at a glance the *P. E.* for a given value of r

[1]The author has found the Gem adding machine serviceable for work in which there is no necessity for printed records such as the Burroughs, Standard, Wales and other high-priced machines afford.

and n, but also determine the value of n, *i. e.*, the number of observations, needed to establish a given degree of correlation with any assigned degree of accuracy. In our illustrative case, since $n = 50$, and $r = .92$, we note that $P. E._r$ is less than 0.0181, that 200 observations would have reduced the error to less than .0091, etc. If our correlation had been lower, say 0.30, the error for 50 cases would have risen to 0.0868.

TABLE 5

Probable Error of r for Various Values of r and of n (*Yule*)

VALUES OF n	$r = 0$	$r = .1$	$r = .2$	$r = .3$	$r = .4$	$r = .5$	$r = .6$	$r = .7$	$r = .8$	$r = .9$	$r = 1$
25	.1349	.1335	.1295	.1228	.1133	.1012	.0863	.0688	.0486	.0256	.0000
50	.0954	.0944	.0916	.0868	.0801	.0715	.0610	.0486	.0343	.0181	.0000
75	.0779	.0771	.0748	.0709	.0654	.0584	.0498	.0397	.0280	.0148	.0000
100	.0674	.0668	.0648	.0614	.0567	.0506	.0432	.0344	.0243	.0128	.0000
200	.0478	.0473	.0459	.0435	.0402	.0359	.0306	.0244	.0172	.0091	.0000
300	.0389	.0386	.0374	.0354	.0327	.0292	.0249	.0199	.0140	.0074	.0000
400	.0337	.0334	.0324	.0307	.0283	.0253	.0216	.0172	.0121	.0064	.0000
500	.0302	.0299	.0290	.0274	.0253	.0226	.0193	.0154	.0109	.0057	.0000
600	.0275	.0273	.0264	.0251	.0231	.0207	.0176	.0140	.0099	.0052	.0000
700	.0255	.0252	.0245	.0232	.0214	.0191	.0163	.0130	.0092	.0048	.0000
800	.0238	.0236	.0229	.0217	.0200	.0179	.0153	.0122	.0086	.0045	.0000
900	.0225	.0223	.0216	.0205	.0189	.0169	.0144	.0115	.0081	.0043	.0000
1000	.0213	.0211	.0205	.0194	.0179	.0160	.0137	.0109	.0077	.0041	.0000

In general, a correlation, like any other determination, to have claim to scientific attention must be at least twice as large as its *P. E.*, and to be perfectly satisfactory, should be perhaps four to five times as large.

Since, in our illustration, r is some 51 times as large as its *P. E.*, its appearance by mere chance is practically zero and its reliability is practically absolute.

(b) The Method of Rank-Differences (Co-ordination)

The product-moments method gives full and exact weight to the d of each m from the M. If we disregard the magnitude of these d's, however, and regard only the relative order, or station, of individual m's in each array, we may yet measure correlation (co-ordination) by what is known as the method of rank-differences.

For this method, the formula, as developed by Spearman (25), is

$$\varrho = 1 - \frac{\Sigma D^2}{c}, \tag{26}$$

in which ϱ is the coefficient of co-ordination, D is the numerical difference between each corresponding pair of ranks[1] (not to be confused with d, the deviation from the mean), and in which c is the mean value of ΣD^2 by mere chance. Since

$$c = \frac{n\ (n^2 - 1)}{6}, \tag{27}$$

Formula 26 may be written:

$$\varrho = 1 - \frac{6\ \Sigma D^2}{n\ (n^2 - 1)}. \tag{28}$$

For illustration, note in Table 1, Column 11, the series of D's which are squared and summated in Column 12. Boy No. 30 ranks first (weakest) in the distribution for right-hand grip and first in the order for left-hand grip, hence his $D = 0$. Boy No. 52[2] is 3d in the first, and 9th in the second array, hence for him $D = 6$ and $D^2 = 36$. Since $n = 50$, by Formula 27,

$$c = \frac{50\ (2500 - 1)}{6} = 20{,}791;$$

hence, by Formula 26,

$$\varrho = 1 - \frac{2098}{20.791} = 1 - .10 = .90.$$

[1]In this and the following method, cases of 'ties' for a given rank are preferably divided in such a manner as to keep the total number of ranks equal in the two series. If, for instance, two S's rank 5th, they should both be assigned the rank 5.5 (to replace 5 and 6), or if three S's rank 5th, they should all three be assigned the rank 6 (to replace the 5th, 6th and 7th places in the series). This procedure introduces a certain amount of error, but not usually enough to be deemed serious.

On the applicability of the method in general and the relation of its index to the Pearson index of correlation, see especially Betz, Dürr, Spearman and Pearson (*Draper's Biometric Series*, 4:1907, 35).

[2]To avoid possible confusion, it may be explained that two records were discarded, so that the boys' numbers run two over the fifty.

a result in close accordance with that of the product-moments method and obtained in a small fraction of the time.

The probable error of ϱ in the method of rank-differences is somewhat larger than in the Pearson or 'cross-multiple' method. Its formula is

$$P.E._{\varrho} = 0.706 \frac{1 - \varrho^2}{\sqrt{n}} \tag{29}$$

The chief difficulty with the rank-method is its assumption that the unit of rank is equal throughout the scale, whereas, as is evident, the units in all probability should mean less in the middle portion than in the extreme portions of the scale. If we assume that the actual frequency-distribution of each of the compared variables is of the normal or Gaussian form, Pearson has shown that the correlation of grades[1] is then functionally related to the coefficient of co-ordination, and in such a manner that

$$r = 2 \sin \left(\frac{\pi}{6} \varrho\right) \tag{30}$$

However, this formula affects only slightly (at most two units in the second decimal) the value of ϱ, so that, for most purposes, the correction that it affords may be neglected.

The $P.E.$ of r thus calculated by the method of rank-differences is somewhat larger than in the Pearson, or 'cross-multiple' method. Its formula is the same as that of the coefficient of co-ordination from which it is computed (Formula 29).

(c) Spearman's Correlation 'Foot-Rule,' or R-Method

Spearman's "'foot-rule' for measuring correlation" (25) is another and still simpler method of comparison by rank, the essential features of which are the use of D, the numerical difference of station, in place of D^2, and of only those of the D's that indicate a gain in rank (since the losses must equal the gains).

---

[1]The grade of an individual is not quite identical with his rank, but is measured by the number of individuals above him. See Brown, p. 19, for illustrative diagram.

This method, it is important to note, yields an index, R, that is not identical with the Pearson r, though functionally related thereto, as is explained below.

The formula for Spearman's R is

$$R = 1 - \frac{\Sigma g}{c}, \tag{31}$$

in which g is the numerical gain in rank of an individual in the second, as compared with the first series, and in which c is the mean value of Σg by mere chance.[1]

Since

$$c = \frac{n^2 - 1}{6}, \tag{32}$$

we obtain by substitution:

$$R = 1 - \frac{6 \Sigma g}{n^2 - 1}. \tag{33}$$

For illustration, note in Table 1, Column 10, the series of gains (g), which yield $\Sigma g = 106$. As $n = 50$, by Formula 32, $c = (2500 - 1) \div 6 = 416.5$;
hence, by Formula 31,

$$R = 1 - \frac{106}{416.5} = 1 - .25 = .75$$

To determine whether R has any claim to reliability, one may use the formula:[2]

$$P.E._R = \frac{0.43}{\sqrt{n}}. \tag{34}$$

[1]If the result turns out to be a minus quantity, the correlation is inverse, but the value of R is then to be computed by ranking one of the series in reverse order.

[2]This formula has also been called in question by Pearson. See Spearman's rejoinder (28, 282ff).

To convert R-values into r-values, Spearman has proposed the formula:

$$r = \sin\left[\frac{\pi}{2}R\right]; \tag{35}$$

or, for all cases in which R is less than .50, this may be simplified with little loss of accuracy into:

$$r = 1.5\,R. \tag{36}$$

For the quick and accurate conversion of R into r, Table 6, which is based on Formula 35, may be consulted.

In the correlation under examination, since $R = .75$, $r = .93$, or practically the value obtained by the longer Pearson method.

According to Pearson (*Drapers Biom. Series,* 4: 1907), however, the proper formula for the conversion of R-values into r-values is

$$r = 2\cos\frac{\pi}{3}(1 - R) - 1. \tag{37}$$

TABLE 6

Conversion of R-Values into r-Values, in Accordance with Formula 35

R	r	R	r	R	r	R	r	R	r
.00	.00	.20	.31	.40	.59	.60	.81	.80	.95
.01	.01	.21	.32	.41	.60	.61	.82	.81	.96
.02	.03	.22	.34	.42	.61	.62	.83	.82	.96
.03	.05	.23	.35	.43	.62	.63	.84	.83	.96
.04	.06	.24	.37	.44	.64	.64	.84	.84	.97
.05	.07	.25	.38	.45	.65	.65	.85	.85	.97
.06	.08	.26	.40	.46	.66	.66	.86	.86	.98
.07	.11	.27	.41	.47	.67	.67	.87	.87	.98
.08	.13	.28	.43	.48	.69	.68	.88	.88	.98
.09	.14	.29	.44	.49	.70	.69	.88	.89	.99
.10	.16	.30	.45	.50	.71	.70	.89	.90	.99
.11	.17	.31	.47	.51	.72	.71	.90	.91	.99
.12	.19	.32	.48	.52	.73	.72	.90	.92	.99
.13	.20	.33	.50	.53	.74	.73	.91	.93	.99
.14	.22	.34	.51	.54	.75	.74	.92	.94	1.00
.15	.23	.35	.52	.55	.76	.75	.93	.95	1.00
.16	.25	.36	.54	.56	.77	.76	.93	.96	1.00
.17	.26	.37	.55	.57	.78	.77	.94	.97	1.00
.18	.28	.38	.56	.58	.79	.78	.94	.98	1.00
.19	.29	.39	.57	.59	.80	.79	.95	.99	1.00

(d) Comparison of Distribution in Selected Groups

The following method is sometimes useful as a device for preliminary survey, but when used, as it often has been, for a final expression of correlation, it is inferior to the methods already described.

Distribute the data for both series in order as illustrated in Table 1, and divide them into four or five groups on the basis of equal numbers of cases or of equal amounts of deviation. By inspection it is often possible to determine at this juncture whether there is sufficient evidence of a correlation to justify further calculation. For this inspection we may take the cases found in the first group of the first series and examine their distribution in the groups of the second series. Evidently, in the absence of correlation, these cases would be distributed by chance. Thus, in Table 1, let the two series be divided into 5 groups of 10 measurements each. The 10 measurements in the first group for right-hand grip would tend, by chance alone, to be distributed 2 in each of the 5 groups in the second series, but as a matter of fact they are massed in or near the first group (8 in the 1st, 2 in the 2d), hence there is evidently a high degree of correlation.

The distribution in the second series of the remaining groups of the first series may be similarly tested, though an examination of the first group is commonly sufficient.

If the grouping is made in terms of deviation, the number of measurements found in the several groups will usually be unequal; it is then necessary to calculate the distribution of the various groups of the first series into those of the second. Suppose the two series of Table 1 are each divided into five 70-hectogram groups: the right-hand series will subdivide into groups containing 16, 15, 12, 4 and 3 measurements; the left-hand series into groups containing 12, 17, 12, 6 and 3 measurements, respectively. Take the 16 cases in the first group of the first series: by chance it is clear that 12/50, or 3.84 of them, would fall in the first group of the 2d series, 17/50, or 5.44 of them, would fall in the second group, 3.84 in the third, 1.97 in the fourth, and .98 in the fifth. The actual distribution of the 16 cases into these five groups is 11, 5, 0, 0, 0, as compared with the chance distribution, 3.84, 5.44, 3.84, 1.97, .98.

If now we wish not only to explore the distribution of selected groups tentatively for the presence of relationship, but also to present the evidence of the relationship in compact form, we may, as is often done, prepare a table by comparative

averages. To return again to the relationship of right and left-hand strength of grip, we may by this means secure the following tabular statement.[1]

TABLE 7

Relation of Right and Left-Hand Grip by Group Averages (Whipple)

Group in 1st Series	1st 10	2d 10	3d 10	4th 10	5th 10
Average Right-hand Grip	194.9	229.9	271.5	317.9	402.2
Average Left-hand Grip	183.7	226.3	269.2	291.5	394.4

The data from such a table may be thrown into graphic form very simply: let ordinates represent the values of the one series, abscissas the values of the other, so that the one series is plotted as a function of the other: if, then, the first series ranges upward for values from low to high and the second from left to right for values from low to high, a positive correlation will be indicated by a line running in a southwest-northeast direction, inverse correlation by a line running in a northwest-southeast direction, and zero correlation by a vertical or a horizontal line (depending on which series is plotted on the ordinates). In proportion as the correlation is complete the line assumes an oblique position.

(e) Relationships of Contingency

The use of such a classified distribution as Table 7 leads naturally to a consideration of cases in which the comparison of two sets of data can be undertaken in that form only, be-

[1]For examples of this type of presentation, consult Bagley (1), Binet and Vaschide (3). Though this method has been frequently used, a little reflection will show that it is inferior to the several methods described before, because the advantage of weighing the relation of each individual measure is lost by lumping them into averages, and because, moreover, no coefficient of correlation is computed. If the groups were made more numerous and the data presented in the form of a graph showing the entire course of the relationship, the absence of the coefficient might be less serious.

For suggestions for further treatment of what he terms *Fraktionskorrelation*, see Stern, 305ff.

cause further information is wanting. The relationship between two sets of values that can merely be classified qualitatively and not graded by magnitude or rank is termed a contingency. Thus, in Table 8[1] the problem is to determine

TABLE 8

Contingency Table for Inheritance of Temperament (Heymans and Wiersma)

FATHERS.

		MERRY	MELANCHOLY	ALTERNATING	EVEN	TOTALS
SONS.	Merry	122	8	81	67	278
	Melancholy	10	2	7	10	29
	Alternating	70	9	101	68	248
	Even	58	6	66	45	175
	Totals	260	25	255	190	730

whether sons tend to resemble their fathers in certain temperamental traits. The formulas for computing the coefficient of contingency from classificatory tables of this sort are too complex to be reproduced here[2]: they are designed to calculate the difference between the numbers actually found in different compartments of the table and the numbers to be expected there by mere chance. For Table 8, the resultant coefficient of mean-square contingency, as it is termed by Pearson, amounts to 0.16.

(f) Method of Association (Four-fold Tables)

The measure of a relationship of contingency is, however, relatively simple if the classification be limited to a four-fold table. The amount of correspondence revealed in such a table is termed by Yule the *degree of association.* Very often the four-fold classification pertains to two traits whose presence or absence is ascertainable, but about whose degree nothing can be said, so that the phrase 'correlation of presence and absence' has been applied to this form of contingency.

[1]Taken from Schuster and Elderton, and derived by them from Heymans and Wiersma (14).

[2]See for details, Brown (p. 13), Betz (29f.), Pearson.

In a typical instance, then, we have given

a = number of cases in which both traits are present,
b = number of cases in which the first trait is present and the second absent,
c = number of cases in which the second trait is present and the first absent,
d = number of cases in which both traits are absent.

On the basis of these values, several formulas have been developed by Yule for the coëfficient of association, Q.

The simplest is

$$Q = \frac{ad - bc}{ad + bc}. \tag{38}$$

This formula gives values which are not identical with r, but stand roughly in the relation 4:3 with it. The substitution of $\sqrt{ad}$ and $\sqrt{bc}$ in both numerator and denominator has been recommended more recently by the same author: the value of Q then approaches nearer Spearman's R.

A third formula devised by Yule, which has the practical merit of yielding values comparable with the ordinary r, is

$$r = \sin \frac{\pi}{2} \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}. \tag{39}$$

If, in this formula, we replace the sine by the cosine of its complement, we secure

$$r = \cos \left[\frac{\pi}{2} - \frac{\pi}{2} \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \right]$$

which we can reduce to

$$r = \cos \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \pi. \tag{40}$$

The probable error, provided *ab* is not very unequal to *cd*, may be taken as

$$P.E._r = \frac{1.1}{\sqrt{n}}. \qquad (41)$$

Our example of four-fold classification (Table 9) is drawn by Stern (30, p. 310) from Heymans (13, p. 336). In it we find $a = 14$, $b = 17$, $c = 7$, $d = 72$. Substituting in Formula 40, we have

$$r = \cos \frac{\sqrt{7 \times 17}}{\sqrt{14 \times 72} + \sqrt{7 \times 17}} \pi = \cos 46^\circ = 0.69$$[1]

TABLE 9

Association of Good Judgment with Phlegmatic Temperament (Heymans)

	PHLEGMATIC	NOT PHLEGMATIC	TOTALS
Good Judgment	14	17	31
Not Good Judgment	7	72	79
Totals	21	89	110

Now, this same formula can often be used for preliminary exploration of series in which the degree of presence of a trait is known, and which may, therefore, be treated, if desired, by the more elaborate methods. For this purpose, assume that all measurements greater than M (or the median), *i. e.*, all *plus* cases, signify the presence of the trait, and all *minus* cases its absence. In Table 1 we find, using the median, 22 cases that are plus in both series, 3 that are plus in right and minus in left-hand grip, 3 that are minus in right and plus in left-hand grip, and 22 that are minus in both. Substituting these values in Formula 40, we have

[1]Stern (p. 312) proposes other methods of treating the four-fold classification, which have the merit of revealing the direction of the association. Thus, in the example above, inspection of the table will indicate that the association tends more in one direction than the other: phlegmatic temperament is more apt to imply good judgment than good judgment is apt to imply phlegmatic temperament. According to Stern's formulas, the first association has the strength of 0.53, the second of 0.32.

$$r = \cos \frac{\sqrt{9}}{\sqrt{484} + \sqrt{9}} \pi, \text{ or } \cos 21.6° = .93,$$

which gives again almost the same value as by the standard formula.

(g) The Method of Unlike Signs

Even this procedure may be simplified by substituting for $\sqrt{bc}$ the *percentage* of cases with unlike signs (U), and for $\sqrt{ad}$ the percentage of cases with like signs (L)[1] with the result,

$$r = \cos \frac{U}{L + U} \pi, \tag{42}$$

or, since $L + U$ must always equal 100, and since $\pi = 180°$, this formula may be condensed, if desired, to

$$r = \cos U\, 1.8° \tag{43}$$

Finally, since U must lie between 50 and 0 for positive and between 50 and 100 for inverse correlations, a table may be prepared[2] from which the values of r may be read directly from any integer value of U.

By reference to the paragraph above it will be seen that in Table 1 we have 6 cases of unlike signs in the 50, hence $U = 12$ and $r = 0.93$, as by other methods.

[1]That is, virtually substituting the arithmetical for the geometrical mean.

[2]Reproduced from an earlier article by the author (35), in which the applicability of the method is discussed more fully.

TABLE 10

Corresponding Values of r and U for Formula 43 (Whipple)
If U is greater than 50, first subtract it from 100, then prefix the minus sign to the correlation indicated

U	r	U	r	U	r	U	r	U	r
0	1.000	10	.951	20	.809	30	.587	40	.309
1	.999	11	.941	21	.790	31	.562	41	.279
2	.998	12	.929	22	.770	32	.536	42	.248
3	.995	13	.917	23	.750	33	.509	43	.218
4	.992	14	.904	24	.728	34	.482	44	.187
5	.987	15	.891	25	.707	35	.454	45	.156
6	.982	16	.876	26	.684	36	.426	46	.125
7	.976	17	.860	27	.661	37	.397	47	.094
8	.968	18	.844	28	.637	38	.368	48	.062
9	.960	19	.827	29	.613	39	.338	49	.031

This method should not be used for final determinations of important correlations because the probable error is too large, but it is a useful device for quick examination of a relation.

3. *The Correction of Obtained Correlations to their True Value*

(a) Correction of the Attenuation Produced by Chance Errors

(The Coefficient of Reliability)

The real correspondence between two traits or capacities is not, as has so often been erroneously supposed, necessarily revealed by the determination of a coefficient of correlation, even by the most approved methods and with a probable error that is satisfactorily small. All measurements, as we have noted, are subject to chance errors of observation. In the determination of averages such errors tend to counterbalance one another, so that if the measurements are sufficiently numerous, the obtained M differs from the true M by an inappreciable amount. In the case of correlations, however, these errors[1] are not eliminated by increasing the number of observations, and their pres-

[1]The phrase 'errors of observation' is to be understood in a wide sense to include not only errors arising from technique, instrumentation, etc., but also *chance* shifts in the disposition of subjects in their attitude toward the test, etc.

ence has the effect of *decreasing* the size of the correlation, so that, in so far as these errors are concerned, the 'raw' or obtained correlation is too small, or, to use Spearman's term, the correlation is 'attenuated' by errors which constitute, from this point of view, constant or systematic errors.

This illusory attenuation of the correlation by errors of observation seems, in fact, a principal cause of the contradictory nature of results that have hitherto been obtained; in experiments in which such errors have been very large, a correlation has not appeared, even when present, and has, in consequence, been erroneously denied. The determination of a small correlation, therefore, opens two possibilities; it may indicate actual absence of correspondence, or it may indicate merely the presence of large chance errors of observation. (Krüger and Spearman, p. 55.)

In order to correct the raw and discover the true r, it is imperative to secure at least two independent series of observations. The formula for correction of attenuation, or the 'expanding' formula, as it might be termed, is then applied as follows:

$$AB_t = \frac{M\ (A_1 B_1,\ A_1 B_2,\ A_2 B_1,\ A_2 B_2)}{M\ (A_1 A_2,\ B_1 B_2)} \tag{44}$$

in which

$AB_t =$ the true correlation,
$M =$ the mean,
$A_1 =$ the 1st series of observations of the trait A,
$A_2 =$ the 2d series of observations of the trait A,
$B_1 =$ the 1st series of observations of the trait B,
$B_2 =$ the 2d series of observations of the trait B,
$A_1 B_1 =$ the raw correlation of A_1 and B_1,
$A_1 A_2 =$ the raw correlation of A_1 and A_2, etc.

Thus the numerator is the M of the four possible r's between the measurements of A and the measurements of B, while the denominator is the M of the r of the two A series and the r of the two B series.

Attention should be called to the opportunity afforded by such correlations as A_1A_2 and B_1B_2 to secure an objective indication of the reliability of the tests used to measure A and B. The correlation thus measured between the results of two different applications of the same test upon the same persons has been used, particularly by recent English investigators, like Spearman, Burt, Wyatt, *et al*, as a *coefficient of reliability*. The principle is simple enough. If the outcome of a test is not disturbed by chance or by constant errors, the ranks of the several *S*'s should be

the same at each trial. Constant errors must, of course, be avoided by other precautions. If, however, chance errors are too obtrusively present, this fact will be revealed by a low correlation between A_1 and A_2. In practise a test whose coefficient of reliability is less than .60 or .70 is in need of rectification—improvement of conditions, larger number of observations—or should be discarded. It should be understood that A_1 and A_2 need not be independent series of tests given by different *E*'s at different sittings, but may be made up from the data obtained during a single sitting, though, as a rule, two sets of data are secured and the correlation is calculated between the first half of the first performance added to the last half of the second performance and the last half of the first performance added to the first half of the second performance.

The point is that, in figuring the coefficient of reliability or in using Formula 44 it is essential that the discrepancies between the two series of measurements of the same trait should really be of a 'chance' character. Suppose, for example, that A represents "memory for nonsense syllables" and has been tested by requiring the subjects to memorize a dozen sets of syllables. It would be wrong, then, to constitute A_1 of the first six sets and A_2 of the second six sets, because the latter half-dozen would be affected by a constant factor—that of practise—to an extent different from the first half-dozen. It would be better to constitute A_1 from the odd- and A_2 from the even-numbered tests.

Again, it must be remembered that a large correction by Formula 44 involves a great increase in the probable error. Hence, it is advisable, after carrying out the tests, to calculate the correlation A_1A_2 at once. If this value turns out to be less than the requisite .60 to .70, the testing should be prolonged or improved until at least this amount is reached. This precaution is particularly necessary when the subjects are few in number, because then the probable error is more dangerous.

The above formula holds for ordinary cases, but if one series of observations, say A, should be known to be much more exact and reliable than the other, then the geometrical should be substituted for the arithmetical *M*. In theory the denominator should always be the geometrical *M*, but the arithmetical *M* is virtually as accurate, and for short series even more desirable. For the mathematical demonstration of this and the following formulas, consult Spearman (22). The correction does not entirely eliminate the uncertainty that arises from the use of 'random samples' for investigation; that must be removed, as already intimated, by the use of more extended series.

Simpson (21a, p. 109) complains that the gain in accuracy by the use of this corrective formula is too little for the labor expended, and that the time had better be spent in getting more accurate data in the first place.

(b) Correction of the 'Constriction' or 'Dilation' Produced by Constant Errors

Attenuation is the result of the operation of chance errors—chance in the sense that the deviation of any measurement

takes place independently of the deviation of any other measurement. If, however, some influence is at work which affects all the measurements of one or of both series, such a constant factor or constant error will prove a source of disturbance that may either increase or decrease the obtained correlation. Such disturbances will result from the operation of any factor which is not strictly relevant to the correspondence under examination.

If an irrelevant factor affects both of the series, it is evident that the correlation will be unduly increased or 'dilated.' Suppose, for example, that one wished to determine the correlation of pitch discrimination with the discrimination of lifted weights, and that the subjects of the experiments were of different ages. Then, since the two capacities in question both tend to improve with age, this common dependence on age will clearly tend to induce the appearance of a correlation, even if there really be none between the capacities themselves when compared under uniform conditions of age.[1]

If an irrelevant factor affects but one of the series, it is evident that the correlation will be unduly decreased or 'constricted,' *i. e.*, the irrelevant influence will tend to reduce any proportionality that really exists between the two series. To quote an example from Spearman, a correlation of 0.49 was discovered between pitch discrimination and school standing, but it was likewise discovered that more than half the children had 'taken lessons,' and thus had the opportunity for special training in the observation of pitches.

These constant irrelevant factors may not always be excluded,

---

[1] This undiscovered or neglected influence of age has been a very common source of error in many studies of correlation. Obviously, this particular irrelevancy may be eliminated practically by proper selection of subjects for the investigation, or it may be eliminated by manipulation of the results in various ways besides that here described: see, for example, Bagley.

but their force can frequently be measured and allowed for by the following formula:[1]

$$AB_t = \frac{AB_a - AC \cdot BC}{\sqrt{(1 - AC^2)\ (1 - BC^2)}}, \qquad (45)$$

in which

AB_t = the true correlation between A and B,
AB_a = the apparent correlation between A and B.
AC = the direct correlation between A and any irrelevant factor, C,
BC = the direct correlation between B and C.

If, as is most often the case, the irrelevant factor affects but one series, this influence of 'constriction' may be excluded by the simpler formula:

$$AB_t = \frac{AB_a}{\sqrt{(1 - AC^2)}} \qquad (46)$$

Thus, in the example mentioned, the correlation between pitch discrimination and its disturbing factor, musical training, was found by computation, to be 0.61; hence, by Formula 46,

$$r = \frac{0.49}{\sqrt{1 - 0.61^2}} = 0.62.$$

From the above considerations, it follows that the experimenter must define with some exactness the traits that are to be examined for a possible correlation, and that he must not seek to establish the correlation until, by means of suitable preliminary exploration, he has discovered all the irrelevant

[1] AB, AC and BC must first be 'expanded' by Formula 44. The application of these formulas for the correction of irrelevant factors in correlations was first made in psychology by Spearman. The mathematical processes concerned therein were developed independently and extended further by Yule, whose theory of 'partial coefficients' may be consulted by those who are interested in the mathematics of correlation (38, Ch. 12).

factors that might disturb the correspondence. The mere mechanical computation of an index of correlation does not, then, demonstrate the existence of a real correlation, or at least, does not accurately and certainly define its nature. Hence, while, as we have seen, we may very hopefully look to correlational work for revelation of the functional disposition of mind, this is no royal road to the attainment of that end, but is a road that can itself be entered upon only after a preparatory survey and critical inspection of the problem in hand has afforded sufficient acquaintance with the traits and capacities that are therein concerned. One must be a psychologist as well as a statistician.[1]

4. *Intercorrelations, Correlations of Pooled Results, etc.*

Tendencies toward functional correspondence may obviously exist not only between two variables, A and B, but also between these and other variables, C, D, etc. It is possible to determine the index of various intercorrelations for groups of data like these, but the formulas are too elaborate to be reproduced here in their entirety, and attention will be called only to two relatively simple problems.[2]

The first problem is that which arises when we wish to 'pool' the results of two or more tests of the same individuals for the purpose of comparing the resultant values with other measurements, as, for example, when we have subjected a group of school children to two or more different mental tests and wish to see whether their average efficiency in these tests correlates with their school standing. It is not permissible to compute the simple correlations and take their average, for the average of several correlations is not the same as their average correlation. Nor is it feasible simply to add together the respective

---

[1]The study of the functional correlation of five well-known tests by Krüger and Spearman affords an admirable illustration of the value of such a combination of sound psychology and sound statistics. Note, however, the criticism of Stern (30, p. 304), though Stern's formula erroneously contains a radical in the numerator.

[2]For further discussion of multiple correlation, consult especially Brown (4, 9-11), Betz (2, 31-33), Yule and other treatises on correlation.

standings of the individuals in the different tests, since the data are often incommensurate (seconds, inches, grams, etc.), while, if we resort to rank order, we deliberately throw away the more exact measures that the original magnitudes obtained by testing have placed at our disposal. Again, the resort to the use of percents of the average performance as a means of indicating the respective standings in the different tests implies that the variability of any series of measures is proportionate to the average of the series, which is an unwarranted assumption.

To meet these difficulties Woodworth (37) has proposed to regard the average of any test as zero. The unit of deviation from this zero is taken directly from the measure of variability *S. D.* (or possibly *A. D.*) of the given array, and the standings of individuals are then computed as deviations, plus or minus, of so and so many times this unit. The arithmetic, then, is simple. Divide each individual deviation (d), as afforded by the original figures, by the *S. D.* of the series and use the resulting quotient (which we may term the 'reduced measure') to represent the individual's standing for subsequent combination or comparison with other tests similarly treated.[1]

By appropriate substitution in the Pearson formula for correlation, it can be shown that if a_1, a_2, a_3, etc., are the reduced measures for individuals 1, 2, 3, etc., in Test A, and b_1, b_2, b_3, etc., are the corresponding values in Test B, then the correlation of A and B is given by the simple formula

$$r = \frac{\Sigma\, ab}{n} \tag{47}$$

In other words, when we have computed the reduced measures of two arrays, the coefficient of correlation between them is the average of the products of the various reduced measures.

[1]Compare the suggestion of Weiss (34): regard the average in any test as 50, divide 50 by the average to determine the value of one unit in the test and multiply this value into the actual number of units accomplished by the person tested. Use the resultant product to represent the new rank of the individual in terms of the average of the group. This method avoids minus quantities, but would appear to be less useful for further computations than that proposed by Woodworth.

Again, the average position (p) of an individual in two tests, A and B, is obviously the sum of his two reduced measures, a and b, divided by 2. If, for other purposes, we have been led to compute the p of each individual, we can then find r by a still shorter method, viz., square the several p's, take the average, multiply it by 2 and subtract 1 from the product.

$$r = 2 \text{ Aver. } p^2 - 1. \tag{48}$$

Formula 48 is really the simplest case falling under the following more general formula.[1] Let t = the number of tests, to find the average correlation among t tests,

$$\text{Av } r = \frac{t \text{ Aver } p^2 - 1}{}. \tag{49}$$

Finally, as a second problem of intercorrelation, mention should be made of another method of dealing with intercorrelations, which has in it much of psychological interest. If three or more psychological traits show intercorrelations one with another, the question may be raised as to whether the intercorrelations are not due to the presence of some common factor to which all the capacities are functionally related, whether, in other words, these correlations may not arise from a single underlying cause. If such a common or 'central' factor be assumed to be present, we may test the validity of the assumption by mathematical procedure, leaving the exact nature of the factor out of consideration for the time being. For example, if for any given capacity, A, we have obtained two independent measurements, A_1 and A_2, and if for two other capacities, we have obtained the measurements B and C respectively, then the correlation (AF) between the capacity A and the hypothetical common or central factor, F, may be determined by the formula:

[1]For further formulas based upon this method and for a discussion of the advantages and disadvantages of using *A. D.* in place of *S. D.* as the unit of variability in calculating reduced measures, consult Woodworth (37). In Spearman (29) will be found other formulas for computing correlations of sums or differences.

$$AF = \frac{M\ (AB,\ AC)}{M\ (A_1A_2,\ BC)}, \tag{50}$$

in which A B represents the direct correlation of A and B, and the other paired symbols have like meanings, while M is the mean (15, p. 88).

In illustration, Krüger and Spearman found the following values—correlation of pitch discrimination with the Ebbinghaus completion test, 0.65, with adding, 0.66, correlation of two measurements of pitch discrimination, 0.87, correlation of the Ebbinghaus test with adding 0.71; hence the correlation of pitch discrimination with the hypothetical central factor is the M of 0.65 and 0.66 divided by the M of 0.87 and 0.71, or 0.83.[1]

REFERENCES[2]

(1) W. C. Bagley, On the correlation of mental and motor ability in school children. *AmJPs*, 12: 1901, 193-205.

(2) W. Betz, Ueber Korrelation. Beiheft zu *ZAngPs*, Leipzig, 1911. Pp. 88. (With bibliography of over 100 titles.)

(3) A. Binet and N. Vaschide, Expériences de force musculaire et de fond chez les jeunes garçons, *AnPs*, 4: 1897 (1898), 15-63.

(4) W. Brown, The use of the theory of correlation in psychology. Cambridge, Eng., 1910. Pp. 83. (Section C published separately, *BrJPs*, 3: 1910, 296-322.)

(5) C. B. Davenport. Statistical methods, with special reference to biological variation. New York, 1899. Pp. 148. (2d revised edition, 1904.)

(6) E. Davenport, Principles of breeding; a treatise on thremmatology. Boston, 1907. Pp. 713. (With an appendix on statistical methods by H. Rietz.)

(7) K. Dunlap, Obtaining the mean variation with the aid of a calculating machine. *PsR*, 20: 1913, 154-7.

(8) E. Dürr, Review of Spearman (22, 23), *ZPs*, 41: 1906, 450. Also, Erwiderung, *ZPs*, 42: 1906, 470-2.

(9) W. P. Elderton, Frequency curves and correlation. London, 1906.

(10) F. Galton, Correlations and their measurement, chiefly from anthropometric data. *Proc. Royal Soc.* London, 45: 1888, 135.

(11) F. Galton, Natural inheritance. London, 1889. Pp. 259.

---

[1]This central factor is tentatively ascribed by these authors to some psychophysiological condition, possibly general neural plasticity. In another article Spearman speaks of "General Discrimination" or "General Intelligence," or the "Intellectual Function," and again (12) of "General Ability." Here, again, is revealed the instructive wealth of possibilities in the application of adequate methodological treatment to psychic life.

[2]Consult the *List of Abbreviations* for the exact titles of periodicals in these and subsequent references.

(12) B. Hart and C. Spearman, General ability, its existence and nature. *BrJPs*, 5: 1912, 51-79.
(13) G. Heymans, Ueber einige psychische Korrelationen. *ZAngPs*, 1: 1908, 313-383.
(14) G. Heymans und E. Wiersma, Beiträge zur speziellen Psychologie auf Grund einer Massenuntersuchung. *ZPs*, 42: 1906, 81-127, 258-301; 43: 1906, 321-373; 45: 1907, 321-333; 51: 1909, 1-72.
(15) F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. *ZPs*, 44: 1907, 50-114.
(16) A. Lehmann and R. Pedersen, Das Wetter und unsere Arbeit. *ArGsPs*, 10: 1907, 1-104, especially 15-26.
(17) M. Merriman, A textbook on the method of least squares. New York, 1884. Pp. 194.
(18) K. Pearson, Mathematical contributions to the theory of evolution. (A series of papers.) *PhTraRoSoc*, 1895 on. (See also *Drapers Company Research Memoirs, Biometric Series*, London, and *Biometrica*, various articles).
(19) H. Rietz: see E. Davenport.
(20) E. C. Sanford, A course in experimental psychology. Boston, 1895 and 1898. Pp. 449.
(21) E. W. Scripture, The new psychology. London, 1897. Pp. 500.
(21a) B. R. Simpson, Correlations of mental abilities. (*Columbia ConEdTSer*, No. 53.) New York, 1912.
(22) C. Spearman, The proof and measurement of association between two things. *AmJPs*, 15: 1904, 72-101.
(23) C. Spearman, General intelligence objectively determined and measured. *AmJPs*, 15: 1904, 201-293.
(24) C. Spearman, Proof and disproof of correlation. *AmJPs*, 16: 1905, 228-231.
(25) C. Spearman, 'Footrule' for measuring correlation. *BrJPs*, 2: 1906, 89-109.
(26) C. Spearman, Entgegnung. *ZPs*, 42: 1906, 467-470.
(27) C. Spearman, Demonstration of formulæ for true measurements of correlation. *AmJPs*, 18: 1907, 161-9.
(28) C. Spearman, Correlation calculated from faulty data. *BrJPs*, 3: 1910, 271-295.
(29) C. Spearman, Correlations of sums or differences. *BrJPs*, 5: 1913, 417-426.
(30) W. Stern, Die differentielle Psychologie. Leipzig, 1911. Especially chs. x to xx. (Consult also for bibliography.)
(31) E. L. Thorndike, Educational psychology. New York, 2d ed., 1910. Pp. 248.
(32) E. L. Thorndike, An introduction to the theory of mental and social measurements. New York, 2d ed., 1913. Pp. 271.
(33) E. B. Titchener, Experimental psychology. Vol. ii, Quantitative experiments, Parts I and II. New York, 1905.
(34) A. P. Weiss. On methods of mental measurement, especially in school and college. *JEdPs*, 2: 1911, 555-563.
(35) G. M. Whipple, A quick method for determining the index of correlation. *AmJPs*, 18: 1907, 322-5.
(36) C. Wissler, The correlation of mental and physical tests. *PsMon*, 3: No. 6, 1901. Pp. 62.
(37) R. S. Woodworth, Combining the results of several tests: a study in statistical method. *PsR*, 19: 1912, 97-123.
(38) G. U. Yule, An introduction to the theory of statistics. London, 2d edition, 1912. Pp. 381.

THE TESTS

CHAPTER IV

ANTHROPOMETRIC TESTS

The tests embraced in this chapter have been developed primarily as anthropometric tests. They do not include tests of physical capacity or function (Chapter V), but simply measurements of bodily size or dimension.

The number of such measurements that have been made and recorded runs well into the hundreds, and an extensive literature has appeared. The science of anthropometry has developed partly in connection with anthropology and sociology, partly in connection with the study of physical development, including bodily growth, hygiene, gymnastic and athletic training. In recent years, moreover, a not inconsiderable contribution has been made by psychologists, physicians, educators and other investigators who have been interested in the correlation between bodily and mental traits.

It is this last-mentioned phase of anthropometry that concerns us, and hence only a few important measurements that have assumed special importance in conjunction with other physical and with mental tests are here considered.

The references which follow will enable the reader to study the development of anthropometry and the application of anthropometric tests at large. Bertillon and Galton should be consulted by those who are interested in the use of anthropometric measurements in the identification of criminals: Key and Hertel have given special consideration to the relation of growth to disease and to hygienic conditions. Anthropometric charts or record-books have been published by E. Hitchcock, D. A. Sargent, J. W. Seaver, W. W. Hastings, Anna Wood, L. H. Gulick and others.

REFERENCES

(1) A. Bertillon, Signaletic instructions; including the theory and practice of anthropometrical identification. Eng. Trans., Chicago and New York. 1896. Pp. 260 + 81 plates, charts, etc.

(2) H. G. Beyer, The growth of United States naval cadets. *ProcUSNInst*, 21: No. 2, 1895.

(3) F. Boas, The correlation of anatomical or physiological measurements. *AmAnt*, 4: 1894, 313.

(4) F. Boas, The growth of children. *Sci.*, 20: 1892, 351-352.

(5) F. Boas. On Dr. W. T. Porter's investigation of the growth of the school children of St. Louis. *Sci.*, 1: 1895, 225.

(6) F. Boas, Growth of first-born children. *Sci.*, 1: 1895, 402.

(7) F. Boas, Growth of Toronto children. *RepBrAAdSci* for 1897, 1898, 443.

(8) F. Boas, Growth of American children. *RepComEd*, 1896-7, ii, 1555.

(9) H. P. Bowditch, The growth of children. *10th An. Rep. State Brd. Health*, Mass., 1879, 35-62.

(10) H. P. Bowditch, The growth of children, studied by Galton's method of percentile grades. *22d An. Rep. State Brd. Health*, Mass., 1891, 479.

(11) F. Burk, Growth of children in height and weight. *AmJPs*, 9: 1897-8, 253.

(12) A. Engelsperger und O. Ziegler, Beiträge zur Kenntnis der physischen und psychischen Natur des sechsjährigen in die Schule eintretenden Kindes. Anthropometrisches Teil. *EPd*, 1: 1905, 173-235.

(13) Lucy Ernst und E. Meumann, Das Schulkind in seiner körperlichen und geistigen Entwicklung. Part 1, by L. Ernst, Leipzig, 1906. Pp. 165.

(14) W. Farr, *et al.*, Table showing the relative statures of boys at the age of 11 to 12 years under different social and physical conditions of life. *RepBrAAdSci*, 1880, i, 127.

(15) F. Galton, Proposal to apply for anthropological statistics from schools. *JAntInst*, 3: 1873-4, 308.

(16) F. Galton, On the height and weight of boys aged 14 in town and country schools. *Nature*, 5: 1875-6, 174.

(17) F. Galton, Range in height of males at each age and in several classes. *RepBrAAdSci*, 51: 1882, 250.

(18) F. Galton, Head growth in students at the University of Cambridge. *Nature*, 38: 1888-9, 14.

(19) F. Galton, Useful anthropometry. *ProcAmAAdPhysEd*, 6: 1891, 51.

(20) F. Galton, Anthropometrical instruments. *AntJ*, 16: 2.

(21) F. Galton, Anthropometric percentiles. *Nature*, 31: 1884-5, 223.

(22) F. Galton, Finger-prints. London, 1892. Pp. 216.

(23) A. Gihon, Physical measurements. *Wood's Ref. Hand-bk. Med. Sci.*, 5: 1887, 667.

(24) N. A. Gratsianoff, Data for the study of physical development in childhood (in Russian). St. Petersburg, 1889.

(25) J. M. Greenwood, Heights and weights of children. *Public Health Ass. Rep.*, 1891, Concord, N. H., 17: 1892, 199.

(26) E. Hartwell, Anthropometry in the United States. *Amer. Statis. Ass.*, 3: 554.

(27) E. Hartwell, Rept. of the director of physical training, Boston Normal Schools. *School Doc., No.* 8, Boston, 1894.

(28) E. Hartwell, Rept. of the director of physical training. *School Doc., No.* 4. Boston, 1895.

(29) E. Hartwell, Bowditch's law of growth and what it teaches. Reprint from 10*th Proc. Am. A. Ad. Phys. Ed.*, Concord, N. H., 1896.

(30) W. Hastings, A manual for physical measurements for use in normal schools, public and preparatory schools, etc. Springfield, Mass., 1902. Pp. 112.

(31) A. Hertel, Neuere Untersuchungen über den allgemeinen Gesundheits-Zustand der Schülerinen. *Schulgesundheitspflege*, Nos. 6 and 7, 1888, 167-183, 201-215. (Contains a review of the work of Key.)

(32) E. Hitchcock, Jr., Physical measurements, fallacies and errors. *ProcAmAAdPhysEd*, 1887, 35.

(33) E. Hitchcock, Jr., A synoptic exhibit of 15,000 physical examinations. Ithaca, 1890.

(34) E. Hitchcock, Comparative study of measurements of male and female students at Amherst, Mt. Holyoke, and Wellesley College. *ProcAmAAdPhysEd*, 6: 1891, 37.

(35) E. Hitchcock, The results of anthropometry as derived from the measurements of the students of Amherst College. Amherst, 1892.

(36) E. Hitchcock and H. Seelye, An anthropometric manual, giving the average and mean physical measurements and tests of male college students, and modes of securing them. Amherst, 1889.

(37) B. Holmes, A study of child growth, being a review of the work of Dr. Wm. T. Porter of St. Louis. *N. Y. Med. J.*, 60: 1894, 417.

(38) W. Jackson, Jr., Graphic methods in anthropometry. *Phys. Ed.*, 2: 1893, 89.

(39) A. Key, Schulhygienische Untersuchungen, trans. and edited, somewhat condensed, from the Swedish, by L. Burgerstein, Hamburg and Leipzig, 1899. (Report of the Swedish Commission of 1882 to investigate hygienic conditions of the schools.)

(40) A. MacDonald, Experimental study of children, including anthropometrical and psychophysical measurements of Washington school children. Reprint of chs. 21 and 25 of *RepComEd*, 1897-8, Washington, 1899.

(41) P. Malling-Hansen, Perioden im Gewicht der Kinder und in der Sonnenwärme. Copenhagen, 1886.

(42) S. Moon, Measurements of the boys of the McDonogh School. McDonogh, Md., 1892. Pp. 46.

(43) K. Pearson, Growth of St. Louis children. *Nature*, 51: 1894, 145-6.

(44) G. Peckham, The growth of children. *Rep. Wis. Brd. Health*, 1881, 28.

(45) G. Peckham, Various observations on growth. *Ibid.*, 1882, 185.

(46) W. T. Porter, On the application to individual school children of the mean values derived from anthropological measurements by the generalizing method. *Pub. Amer. Statis. Ass.*, n. s. 3: 1892-3, 576.

(47) W. T. Porter, The growth of St. Louis children. *Trans. Acad. Sci.*, St. Louis, 6: 1894, 263-426.

(48) W. T. Porter, Physical basis of precocity and dullness. *Ibid.*, 6: 1893-4, 161-181.

(49) W. T. Porter, The relation between growth of children and their deviation from the physical type of their sex and age. *Ibid.*, 233-250.

(50) W. T. Porter. Use of anthropometrical measurements in schools. *EdR*, 11: 1896, 126-133.

(51) C. Roberts, A manual of anthropometry.

(52) Sack, Physical development of the children in the middle schools of Moscow, 1892.
(53) D. Sargent, Report on anthropometric measurements. *ProcAm AAdPhysEd*, 2: 1886, 6.
(54) D. Sargent, Anthropometric apparatus, with directions for measuring and testing the principal physical characteristics of the human body. Cambridge, Mass., 1887.
(55) D. Sargent, The physical proportions of the typical man. *Scribner's Mag.*, 2: 1887, 3-17.
(55a) D. Sargent, The physical development of women. *Ibid.*, 5: 1889, 541.
(56) D. Sargent, Strength-tests and strong men at Harvard. *J. Boston Soc. Med. Sci.*, No. 13, 1896-7.
(57) G. Schultz, Some new anthropometrical data. *Yale Med. J.*, 2: 1895-6, 149.
(58) J. Seaver, Anthropometry and physical examination. New Haven, 1896. Pp. 200.
(59) G. Sergi, An anthropological cabinet for pedagogic purposes. *Ed.*, 7: 1886, 42-9.
(60) C. Stratz, Der Körper des Kindes, für Eltern, Erzieher, Aerzte, u. Künstler. Stuttgart, 1904.
(61) F. Swain, Anthropometric measurements. *ProcAmAAdPhysEd*, 3: 1887, 43.
(62) G. Tarbell, On the height, weight and relative growth of normal and feeble-minded children. *Proc. 6th An. Session Ass. Med. Officers Amer. Inst. Idiotic and Feeble-minded Persons.* Phila., 1883, 188.
(63) R. Thoma, Untersuchungen über die Grösse u. das Gewicht der anatomischen Bestandtheile des menschlichen Körpers. 1882.
(64) E. B. Titchener, Anthropometry and experimental psychology. *PhR*, 2: 1893, 187-192.
(65) F. Tuckermann, Anthropometric data based upon nearly 3000 measurements taken from students. Amherst, 1888.
(66) G. West, Worcester school children; the growth of the body, head and face. *Sci.*, 21: 1893, 2-4.
(67) G. West, Observations on the relation of physical development to intellectual ability, made on the school children of Toronto, Canada. *Sci.*, n. s. 4: 1896, 156.
(68) C. Wissler, Growth of boys; correlations for the annual increments. *AmAnt*, n. s. 5: 1903, 81.
(69) M. Wood, Anthropometric tables, arranged after the method of percentile grades, of the measurements of 1500 Wellesley students.
(70) M. Wood, Anthropometric tables, compiled from the measurements of 1100 Wellesley College students. 1890.
(71) M. Wood, Statistical tables concerning the class of 1891 of Wellesley College, numbering 104 women.
(72) M. Wood, Statistical tables, showing certain measurements of forty freshmen of Wellesley College (before and after gymnasium training). 1892.

TEST I

Height, standing and sitting.—The general purpose of this test is, of course, to furnish a measurement of height as an index of physical size or growth for the sake of comparison with mental traits or with other physical traits. It is included in practically every series of tests that include any physical measurements.

APPARATUS.—Stadiometer (Fig. 4). Small calipers (Fig. 5) or millimeter rule.

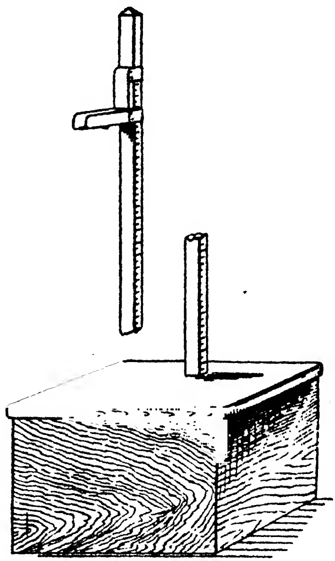


FIG. 4. STADIOMETER, OR HEIGHT STAND.
Graduated in tenths of inches on one side and millimeters on the other.

METHOD.—(1) For standing height, the examiner, *E*, should, when feasible, have the subject, *S*, remove his shoes, and stand on the stadiometer with the heels together and with heels, buttocks and spine between the shoulders, and the head, all in contact with the measuring rod. The chin must not be unduly raised or depressed. *E* then brings down the sliding arm of the instrument until it rests squarely, but without excessive pressure, upon *S*'s head.

(2) For sitting height, let *S* sit erect upon the stand of the stadiometer with spine and head in contact with the measuring rod.

RESULTS.—(1) The best norms of stature are doubtless those calculated by Boas (4)[1] from studies by various investigators of school children (45,151 boys and 43,298 girls) in Boston, St. Louis, Milwaukee, Toronto and Oakland, Cal.[2] For the sake of comparison with these norms and with the norms for strength

FIG. 5. VERNIER CALIPER, FOR EXTERNAL, INTERNAL AND DEPTH MEASURING. Fitted with both English and metric scales and verniers for each, reading to 1-128 of an inch and 1-10 of a millimeter.

of grip, vital capacity, etc., to be quoted later, there are given herewith the norms of standing and sitting height derived from the measurement of 2788 boys and 3471 girls by Director Smedley of the Department of Child-Study and Pedagogic Investigation, Chicago (25).

TABLE 11

Norms of Stature of American Children, in cm. (Boas)[3]

Age	5.5	6.5	7.5	8.5	9.5	10.5	11.5
Boys	105.90	111.58	116.83	122.04	126.91	131.78	136.20
Girls	104.88	110.08	116.08	121.21	126.14	131.27	**136.62**
Age	12.5	13.5	14.5	15.5	16.5	17.5	18.5
Boys	140.74	146.00	152.39	159.72	164.90	168.91	171.07
Girls	**142.52**	**148.69**	**153.50**	156.50	158.03	159.14	

[1]The figures in parentheses following names refer to the reference-numbers at the end of the test in which they occur.

[2]The same averages converted into inches may be found in Burk, while these and other studies are summarized by MacDonald. Consult Boas (4, 1555-6) for table showing the distribution of stature of American boys and girls at each age according to the frequency method.

[3]The figures in black-faced type in Tables 11-15 indicate periods in which the averages for girls exceed those for boys of the same age. The rapid growth of puberty and early adolescence is initiated and terminated earlier in girls than in boys.

TABLE 12

Norms of Standing and Sitting Height, in cm. (Smedley)

AGE	STANDING HEIGHT		SITTING HEIGHT		AGE	STANDING HEIGHT		SITTING HEIGHT	
	BOYS	GIRLS	BOYS	GIRLS		BOYS	GIRLS	BOYS	GIRLS
6.0	110.69	109.66	62.40	61.72	12.5	141.89	144.32	74.70	76.29
6.5	113.25	112.51	63.54	62.90	13.0	145.54	147.68	76.24	77.91
7.0	115.82	115.37	64.67	64.07	13.5	149.09	151.04	77.79	79.54
7.5	118.39	118.22	65.78	65.25	14.0	151.92	153.64	79.21	80.99
8.0	120.93	120.49	66.75	66.34	14.5	154.74	156.24	80.64	82.43
8.5	123.48	122.75	67.72	67.43	15.0	158.07	156.83	82.18	83.21
9.0	126.14	125.24	68.79	68.32	15.5	161.41	157.42	83.68	83.99
9.5	128.80	127.74	69.85	69.21	16.0	164.03	158.30	85.43	84.54
10.0	130.91	130.07	70.56	70.05	16.5	166.65	159.18	87.17	85.09
10.5	133.03	132.41	71.26	70.89	17.0	167.85	159.26	88.16	85.20
11.0	135.11	135.35	72.10	72.23	17.5	169.04	159.34	89.14	85.30
11.5	137.19	138.30	72.93	73.58	18.0	171.23	159.42	90.30	85.51
12.0	139.54	141.31	73.80	74.93	18.5	173.41	159.50	91.46	85.72

From Smedley's data are prepared the charts (Figs. 6 and 7) showing percentiles of height for boys and girls at each age from 4 to 18. These charts show graphically the comparative rate of growth at these different ages, and they are especially valuable for locating a given child in the percentile group to which he belongs by simply following the line of his age across the chart until the point representing his measurement is reached.[1]

From these and other statistics the following important results may be gathered:

(2) There is a period of slower growth in height in boys at 11 years of age, and a similar, though less marked, retardation in girls at nine years of age.

[1]These charts have only an approximate accuracy because of the limited number of cases (300-400 in most years). They are reproduced here partly to show the method of displaying norms which we need to follow for quick and accurate rating of performance in various mental as well as physical tests. Evidently percentile tables of this sort necessitate the gathering of data by standardized tests from large numbers of children of both sexes and various ages. The lack of percentile tables for mental tests is one of the most serious handicaps now existing in their application.

The valuable percentile charts of Porter (20) are worthy of consultation by those particularly interested in anthropometric work.

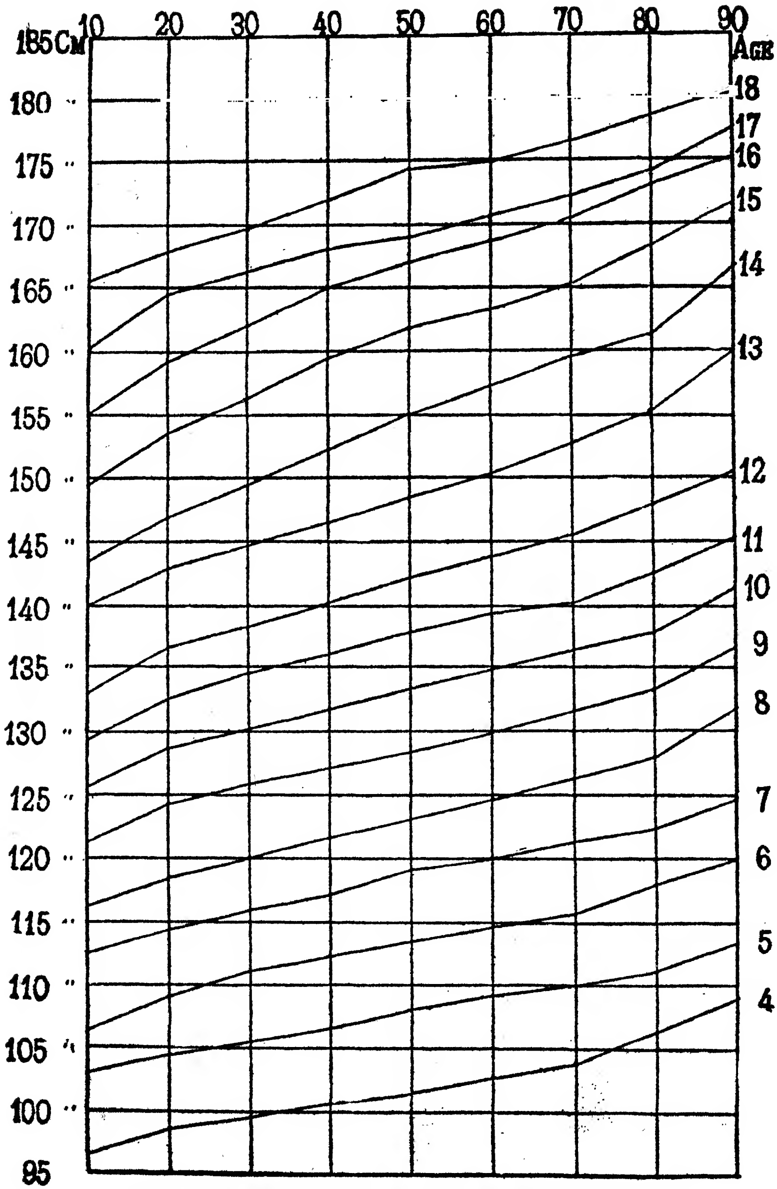


FIG. 6. PERCENTILES OF HEIGHT FOR BOYS (SMEDLEY)

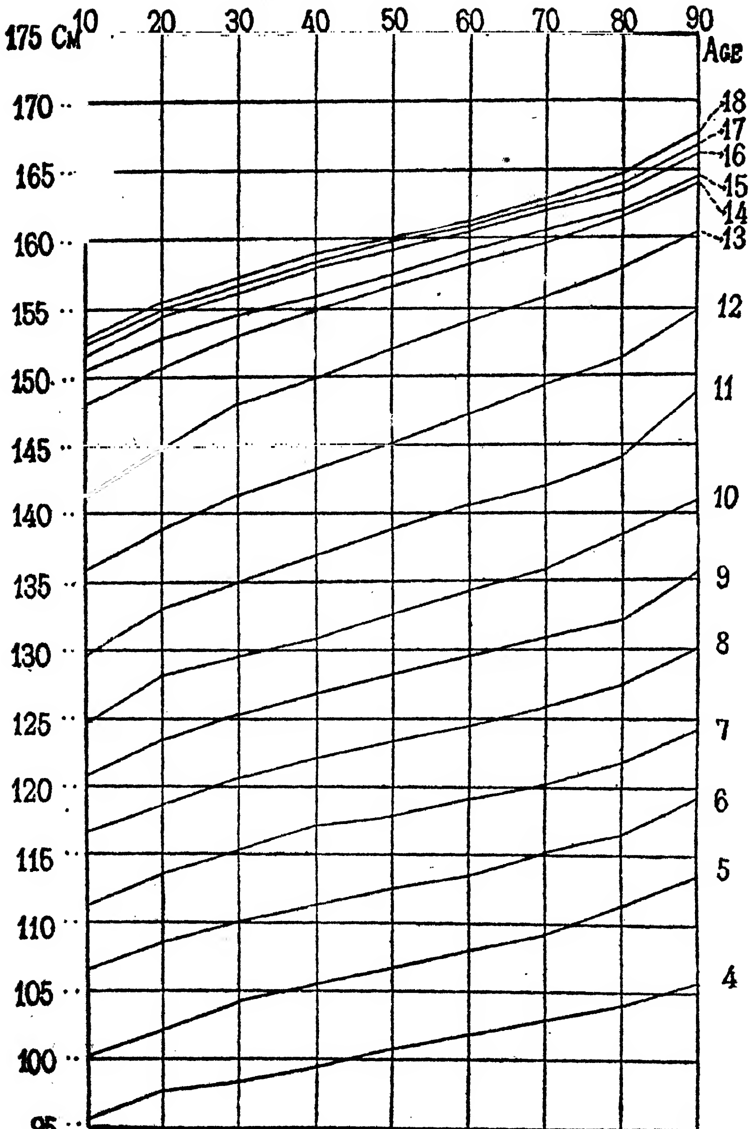


FIG. 7. PERCENTILES OF HEIGHT FOR GIRLS (SMEDLEY)

(3) During the period of approximately 11 to 14 years girls are taller than boys of the same age, because the prepubertal acceleration of growth occurs earlier in girls.

(4) Sitting-height follows the same general laws as standing-height.

(5) Boys continue their growth in height later than do girls, *i. e.*, maturity in height is not reached so early.

(6) Children of purely American descent are taller than those of foreign-born parentage (Bowditch, Peckham).

(7) Race seems to be more important than social or environmental conditions in determining absolute growth, but the latter conditions may influence the rate of growth at different stages of development (Meumann, pp. 82-3).

(8) Children of the non-laboring classes are, as a group, taller than children of the laboring classes (Bowditch, Roberts), and the difference seems to be particularly evident with girls (Pagliani).

(9) First-born are somewhat taller than later-born children, but the amount of the difference is not definitely known (Boas, 4, p. 1596).

(10) According to Bowditch, and also Baldwin, large children make their most rapid growth at an earlier age than small ones, but according to Boas (2, 3) this induction is untenable.

(11) The height of American-born children is modified by density of population. Urban life decreases stature from five years of age upward (Peckham). A similar conclusion is reached by Schmidt.

(12) According to Kline, boys in the public schools are taller than boys in truant schools, save at the age of ten. Similarly, Smedley (27) found that from the tenth year up the boys in the Chicago School for Incorrigibles and Truants were shorter than normal boys.

(13) Gratsianoff and Sack in Russia, and Porter, MacDonald and Smedley in America, have concluded that bright children are taller than dull children, and DeBusk's preliminary study in Colorado points in the same direction. West, however, found exactly the opposite to be true, while Gilbert (11, 12) found no constant relation between height and mental abil-

ity. Porter, DeBusk and Smedley determined mental ability by the relation of grade and age, Gilbert and MacDonald by the teacher's estimate.

The apparent correlation between height and mental ability raises an important question which reappears whenever we discuss the correlation between any physical trait, *e. g.*, weight, strength, vital capacity, etc. and mental ability. The trend of evidence is to the effect that all such correlations, where found, are largely explicable as phenomena of growth, *i. e.*, as correlations with relative maturity. (Cf. Boas, 3; Wissler). This makes intelligible the fact that, in general, the positiveness of all such correlations lessens with age, and that many of them, indeed, become difficult or impossible of demonstration in adults. Thus, to take the correlation in question, a positive correlation is not, of course, to be interpreted as meaning that, taken individually, all tall boys, are, *ipso facto*, bright boys, but that, taken collectively, those boys whose physical condition is good, whose growth is unimpaired by ill-health, faulty nutrition, etc., and who realize to the full the possibility of physical development inherent in them (whether they will ultimately be short or tall) will be found to exhibit the best mental condition and the most rapid mental development. The assertion made by Crampton that the correlations found by Smedley and others between school grade and height, weight, strength, etc., are all due to the earlier pubescence of certain pupils would seem to be negatived by the fact that the relations in question were found in grades before pubescence could have come for any of the pupils in them. Nevertheless, there is probably some truth in Crampton's statement that "statistics for groups or individuals respecting weight, height, strength, mental or physical endurance, medical or social conditions, that are not referred to physiological age are inconsequential and misleading." In general, the whole subject of physiological age and its relations with chronological, psychological and pedagogical age is much in need of more careful and extensive investigation (cf. the author's comment, *JEdPs*, 3: 1912, 410.)

(14) From statistics gathered from 19 institutions, and including 5800 boys and 4800 girls, Goddard finds that the feeble-minded are shorter than normal children, especially at the upper ages, save that moron girls are taller than normal girls from 7 to 9 years. Furthermore, the lower the mental grade, the greater the divergence from the normal rate of growth. Sex differences of stature are less marked in low-grade defectives.

(15) Children with abnormalities are inferior in height to children in general (MacDonald).

NOTES.—The upright measuring rod should be braced in such a manner that it will not be bent out of place by the pressure of *S*'s back. Many *S*'s will be inclined to assume an unnatural position in this examination, especially to stretch themselves:

the apparent height may be increased by as much as 20 to 30 mm. in this way.

If it is not practicable to remove the shoes, height may be taken with them on, and the height of the heel may subsequently be determined by the use of the small calipers or millimeter rule, and then subtracted from the gross height; the resulting error will be very small.

Height, as is well known, decreases slightly during the day, owing to the packing of the intervertebral cartilages and the loss of muscular tone: this loss in height during the day amounts, in the case of young men, to from 10 to 18 mm. It is therefore desirable, for accurate work, to take height measurements at approximately the same period of the day. It might be possible to work out empirically a corrective formula.

REFERENCES

(1) B. T. Baldwin, Individual differences in the correlation of physical growth of elementary and high-school pupils. *JEdPs*, 2: 1911, 150-2.

(2) F. Boas, The growth of children. *Sci.*, 20: 1892, 351-352.

(3) F. Boas, On Dr. W. T. Porter's investigation of the growth of the school children of St. Louis. *Sci.*, n. s. 1: 1895, 225.

(4) F. Boas, The growth of Toronto children. *RepComEd*, 1896-7. ii, 1541-1599.

(5) H. P. Bowditch, The growth of children. *8th An. Rep. State Board Health Mass.*, 1877.

(6) H. P. Bowditch, The growth of children, studied by Galton's method of percentile grades. *22d An. Rep. State Brd. Health Mass.*, 1891, 479.

(7) F. Burk, Growth of children in height and weight. *AmJPs*, 9: 1897-8, 253..

(8) W. G. Chambers, Individual differences in grammar-grade children. *JEdP*, 1: 1910, 61-75.

(8a) C. W. Crampton, The influence of physiological age on scholarship. *PsCl*, 1: 1907, 115-120.

(9) B. W. DeBusk, Height, weight, vital capacity and retardation. *PdSe*, 20: 1913, 89-92.

(10) Lucy Hoesch Ernst, Das Schulkind in seiner körperlichen und geistigen Entwicklung. I Teil, Leipzig, 1906. (Consult ch. 1 for stature of Zurich school children and comparative charts summarizing many other investigations.)

(11) J. A. Gilbert, Researches on the mental and physical development of school children. *SdYalePsLab*, 2: 1894, 40-100.

(12) J. A. Gilbert, Researches upon school children and college students. *UnIowaSdPs*, 1: 1897, 1-39.

(12a) H. H. Goddard, The Height and weight of feeble-minded children in American institutions. *JNeMeDis*, 39: 1912, 217-235.

(13) A. Key (a) Die Pubertätsentwicklung und das Verhältnis der

selben zu den Krankheitserscheinungen der Schuljugend. Sonderabdruck aus den Verhandlungen des 10 *InCgMd.*, 1890; (b) Schulhygienische Untersuchungen. Hamburg u. Leipzig, 1889.
(14) L. W. Kline, Truancy as related to the migratory instinct. *PdSe*, 5: 1898, 381-420.
(15) A. MacDonald, Experimental study of school children, etc. *RepComEd*, 1899 (chs. 21 and 25.)
(16) E. Meumann. Vorlesungen zur Einführung in die exp. Pädagogik. Bd. I. (2d ed.). Leipzig, 1911.
(17) Pagliani, Lo sviluppo umano per atà, sesso conditione sociale ed ethica, Milano. Livelli, 1879.
(18) G. Peckham, The growth of children. *Rep. Wis. Brd. Health*, 1881. 28.
(19) G. Peckham, Various observations on growth. *Ibid.*, 1882, 185.
(20) W. T. Porter, The growth of St. Louis children. *Trans. Acad. Sci.*, St. Louis, 6: 1894, 263-426.
(21) W. T. Porter, Physical basis of precocity and dullness. *Ibid.*, 6: 1893-4, 161-181.
(22) C. Roberts, A manual of anthropometry. London, 1878.
(23) N. Sack, Physical development of the children in the middle schools of Moscow., 1892.
(24) E. Schmidt, Körpergrösse u. Gewicht der Schulkinder des Kreises Saalfeld (Meiningen). 1892.
(25) E. Schuster. First results from the Oxford anthropometric laboratory. *Biometrika*, 8: 1911, 40-51.
(26) F. W. Smedley, Rept. of the dept. of child-study and pedagogic investigation. Reprint from 46th *An. Rep. Brd. Educ.*, Chicago, 1899-1900. Also *RepComEd*, 1902, i., 1095-1115.
(27) F. W. Smedley, do., No. 3, 1900-1901; also 1902, i., 1115-1138.
(28) G. West. Observations on the relation of physical development to intellectual ability made on the school children of Toronto, Canada. *Sci.*, n. s., 4: 1896, 156.
(29) C. Wissler. The correlation of mental and physical tests. *PsMon*, 3: No. 6, 1901. Pp. 62.

TEST 2

Weight.—The general purpose of determining weight is similar to that of determining height, viz.: to furnish an index of physical size or growth as a basis for correlation with other tests or observations.

Apparatus.—Accurate scales, preferably of the type especially devised for anthropometric work, which allow readings to be rapidly and accurately taken in the metric system, with units of 50 g. or twentieths of a kilogram (Fig. 8). If avoirdupois scales are used, they should be divided into tenths of pounds rather than into ounces.

Method and Treatment of Results.—For accurate measurements, weight should be taken without clothes. Where this is

impracticable, the weight of the clothes may be deducted by subsequent measurement. For some comparative purposes, however, the weight of the clothes may be neglected and the figures obtained from the gross weight may be taken for computation, or these figures, better yet, may be corrected by arithmetical computation based upon the weights of the clothes of a limited number of *S*'s.

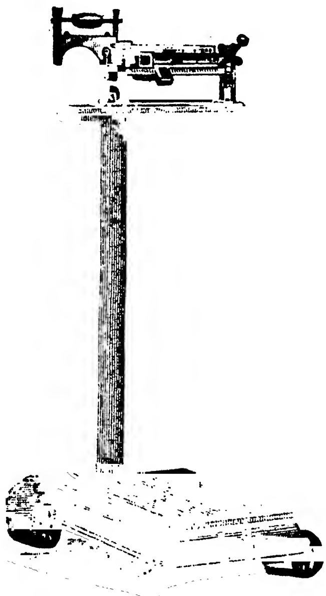


FIG 8. ANTHROPOMETRIC SCALES.

One side of the beams is graduated metric to 100 kilos, by 50-gram divisions, and the other side avoirdupois to 200 pounds, by tenths of a pound.

We may form a tolerably accurate notion of the 'clothing error' by reference to investigations upon this point. Thus, according to W. S. Christopher, who ascertained the weight of the ordinary schoolroom clothing of 121 Chicago children, chiefly in the month of May, "the average weight of the clothing of all the pupils was 5.5 per cent. of the gross weight" (boys, 5.8 per cent.; girls, 5.2 per cent.). These figures varied little with age: obese children wore clothing lighter in proportion to their weight than that worn by others, while "the most variable

element in the clothing was found to be the shoes, especially the shoes worn by the boys." Only a few children wear clothing that weighs more than 7 per cent., or less than 4 per cent. of their gross weight.

RESULTS.—(1) From the data of about 68,000 children in the cities of Boston, St. Louis and Milwaukee Burk derives the norms reproduced in Table 13: the Chicago norms are reproduced in Table 14. For charts of percentile distribution like Figs. 6 and 7 consult Smedley.[1] Means for English university students will be found in Schuster.

TABLE 13

Norms of Weight, in kg. (Burk)

Approx. Age	6.5	7.5	8.5	9,5	10.5	11.5	12.5	13.5	.5	15.5	16.5	17.5
Boys	20.50	22.45	24.72	27.03	29.66	32.07	34.88	38.46	43.18	48.72	54.88	—
Girls	19.69	21.64	23.81	26.03	28.53	31.52	35.70	40.23	44.59	48.40	50.94	52.31

TABLE 14

Norms of Weight, in kg., with Clothing (Smedley)

Age	6.0	7.0	8.0	9.0	10.0	11.0	12.0
Boys	19.738	21.613	23.817	26.336	28.707	31.223	34.151
Girls	18.870	20.974	23.010	25.257	27.795	30.662	34.373

Age	13.0	14.0	15.0	16.0	17.0	18.0
Boys	38.084	42.696	47.993	53.238	57.384	61.283
Girls	38.974	44.219	48.161	50.652	52.386	52.923

(2) As in the case of height, girls exhibit the prepubertal increase in weight some two years earlier than boys, and are for the years 12 to 15 heavier than boys of the same age.

(3) According to Baldwin, the period of maximal increase

[1]The numerical data for these distributions are reproduced without the charts in *RepComEd*, I. 1902, pp. 1120-8.

in weight comes earlier for boys or girls who are above the median height than for those below it.

(4) Growth in weight, as in height, is subject to some lessening of rate at 9 years for girls and at 11 for boys.

(5) Boys continue to increase in weight after girls have practically attained their maximal normal weight. Girls grow most rapidly from 10 to 15 years, boys from 12 to 17 years.

(6) Mean variations in weight are largest during the period of fastest growth, which shows that not all individuals participate equally or evenly in the rapid growth of adolescence.

(7) First-born children exceed later-born children in weight, at least during the period from 6 to 15 years, though the reverse is true of the weight at birth. The difference is slight, but very regular (Boas).

(8) Children of the non-laboring classes are as a group heavier than children of the laboring classes (Bowditch). Pagliani, in Italy, found the children of the poor especially below standard in weight.

(9) Children of American-born parents are heavier than those of foreign-born parents.

(10) There is, naturally, a general correlation between height and weight. Among Oxford students the correlation proved to be 0.66 (Schuster), among Cambridge students 0.49. Again, it is possible to equate growth in height and growth in weight: roughly, boys may be said to add 0.5 kg. in weight for each cm. added in height (Malling-Hansen, Ernst). However, stature is so far from being the chief conditioning factor in weight that in the tables drawn up by Ernst the lightest relative weight (*i. e.*, grams per cm. of height) was rarely found in the shortest child in each group.

(11) The correlation between weight and mental ability or precocity is found to be positive by some investigators, negative by others, and indifferent by still others. Thus, Porter asserts very positively that "precocious children are heavier and dull children lighter than the mean child of the same age," and draws a further practical conclusion that "no child whose weight is below the average for its age should be permitted to enter a school grade beyond the average of its age, except after

such a physical examination as shall make it probable that the child's strength be equal to the strain." Porter's conclusion is confirmed by Smedley at Chicago, and, so far as his limited data suffice, by DeBusk. On the basis of the teacher's estimate of mental ability, Gilbert (8, 9), however, finds no constant relation between weight and such ability, save that from 10 to 14 years the dull children are much heavier than the bright, while West, who used a similar basis, finds a negative correlation throughout.[1]

(12) Both Kline and Smedley find the mean weight of boys in truant schools to be less than that of boys in the public schools, save at the age of 10.

(13) Porter concludes that the acceleration in weight preceding puberty takes place at the same age in dull, mediocre and precocious children, but investigations in New York City seem to oppose this conclusion and indicate rather that puberty and pubertal growth is distinctly earlier in precocious children, *i. e.*, that mental and physical precocity go hand in hand.

(14) From data from 19 institutions, including about 5800 boys and about 4800 girls, Goddard finds that feeble-minded children are distinctly heavier than normal children at birth, but below them at the age of 6, and decidedly below them at later ages, especially from 15 onward, save that moron girls are consistently heavier than normal girls. Moreover, the greater the mental defect, the greater, on the average, is the disparity between the weight of the feeble-minded and the weight of normal children. The feeble-minded also show greater variability in weight than do normals.

(15) Children with abnormalities are below the average in weight (MacDonald).

NOTES.—It is not important to have scales which render possible a very fine measurement, such as fractions of an ounce, because the normal weight of any individual varies from day to day and from hour to hour during the day: the daily variation is, in the case of young men, as high as 0.3 kg. The author, in a long series of observations conducted at the same hour daily,

---

[1]See also the discussion of correlation with mental ability in Test 1.

found gains and losses of more than 1 kg. in 24 hours. It may not be amiss in this connection to point out the absurdity of attaching any significance to small gains or losses that are observed in weighings conducted at occasional and irregular intervals. Severe exercise may reduce the weight by a large amount; *e. g.*, two hours of football practise may take off 2 or 3 kg. from a man who is not yet in training. It is well, however, for comparative purposes, to take weight measurements at approximately the same period of the day.

REFERENCES

(1) B. T. Baldwin. Individual differences in the correlation of physical growth of elementary and high-school pupils. *JEdPs*, 2: 1911, 150-2.

(2) F. Boas, Growth of first-born children. *Sci.*, n. s. 1: 1895, 402-4.

(3) F. Burk, Growth of children in height and weight. *AmJPs*, 9: 1897-8, 253.

(4) W. G. Chambers, Individual differences in grammar-grade children. *JEdPs*, 1: 1910, 61-75.

(5) W. S. Christopher, Report on child-study. Reprint from *An. Rep. Brd. Educ.* Chicago, 1898-1899.

(6) B. W. DeBusk, Height, weight, vital capacity and retardation. *PdSc*, 20: 1913, 89-92.

(7) Lucy Hoesch Ernst, Das Schulkind in seiner körperlichen u. geistigen Entwicklung. I Teil. Leipzig, 1906. Ch. 3.

(8) J. A. Gilbert, Researches on the mental and physical development of school children. *SdYalcPsLab*, 2: 1894, 40-100.

(9) J. A. Gilbert, Researches upon school children and college students. *UnIowaSdPs*, 1: 1897, 1-39.

(9a) H. H. Goddard, The height and weight of feeble-minded children in American institutions. *JNeMcDis*, 39: 1912, 217-235.

(10) L. W. Kline, Truancy as related to the migratory instinct. *PdSc*, 5: 1898, 381-420.

(11) A. MacDonald. Experimental study of school children, etc. Reprint of chs. 21 and 25 of *RepComEd*, 1899.

(12) R. Malling-Hansen (a) Ueber Periodizität im Gewicht der Kinder. Copenhagen, 1883. (b) Perioden im Gewicht der Kinder u. in der Sonnenwärme. Copenhagen, 1886.

(13) W. T. Porter, Physical basis of precocity and dullness. *Trans. Acad. Sci. St. Louis*, 6: 1894, 263-426.

(14) E. Schuster, First results from the Oxford anthropometric laboratory. *Biometrika*, 8: 1911, 40-51.

(15) F. Smedley, Rept. dept. child-study and pedagogic investigation, No. 3, *Rep. Brd. Educ. Chicago*, 1900-1901; also *RepComEd*, 1902, i., 1115-1138.

(16) G. West. Observations on the relation of physical development to intellectual ability, made on the school children of Toronto, Canada. *Sci.*, n. s. 4: 1896, 156.

TEST 3

Diameter of the skull.—This measurement has been commonly conducted for the purpose of investigating the correlation between size of the head and general intelligence. It forms also one of the chief measurements undertaken in the Bertillon system for the identification of criminals. The following directions are adapted from Bertillon's account (1).

A. MEASURING THE LENGTH OF THE HEAD

INSTRUMENT.—Head calipers (Fig. 9).

METHOD.—(1) Seat *S* with his right side toward a window, and stand facing his left side. Hold the left tip of the calipers firmly in place at the glabella (space between the eyebrows),

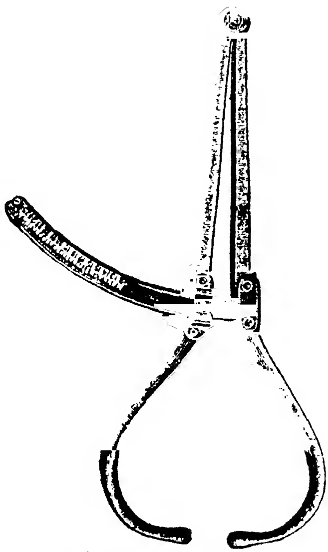


FIG 9. HEAD CALIPERS.

with the tip of the instrument between the thumb and forefinger, and with these resting on the adjacent parts of the forehead to prevent the compass-tip from deviating.

(2) Hold the calipers in an approximately horizontal plane so that the scale is fully lighted by the window, with the right

tip projecting about one cm. beyond the finger-tips of the right hand. Keep the eyes fixed upon the scale; then bring the right tip down over the back and middle of the head until it has passed the most projecting point; then move the tip upward again, making sure that it is well within the hair and in constant contact with the scalp; continue these exploring movements so as to pass the maximal point two or three times, keeping the eyes constantly fixed upon the scale to detect this point.

(3) Remove the calipers and set them by tightening the setscrew at the supposed length; take care to set them accurately within 0.5 mm.

(4) Replace the calipers thus set and tightened, and again execute the exploring movements described in (2). If the setting is correct, the instrument will just *touch* the skin of the head at the maximal point, but will pass over it without undue friction and without necessitating pressure upon its arms: one millimeter too short will produce definite resistance at this point; one millimeter too long, a definite lack of friction. Practise will enable *E* to distinguish the 'feel' of the correctly set instrument, and errors should not exceed 1 mm.

B. MEASURING THE WIDTH OF THE HEAD

INSTRUMENT.—Head calipers as above.

METHOD.—Position of *S*, preliminary exploring movements, setting of the calipers, and subsequent verification follow the same general procedure as in the determination of the length of head. The following additional instructions are to be noted:

(1) *E* stands behind *S*, and is careful to preserve an erect, symmetrical position, in order to ensure equal freedom with both elbows and a symmetrical position of the calipers.

(2) Hold the calipers a short distance from each end; apply the tips first at the upper point of attachment of each ear; then raise them *vertically* and watch the scale to determine the point of greatest width, making several testing movements both upward and downward.

(3) The true maximal diameter in most cases is not yet found, but lies in the same horizontal plane as the preliminary maximal point just determined, and about 3 cm. behind it.

Hence, next move the calipers slowly back and forth two or three times in a *horizontal* plane and determine the true maximal point.

(4) Set the instrument, as in the previous measurement, and verify the setting. In this verification the caliper-points should describe a series of zig-zag movements, in order certainly to traverse the areas of maximal width (usually less than the size of a dime), which might not be traversed if the movements were circular or too coarsely executed.

TREATMENT OF RESULTS. — From the measurements of the length and width of head, the *cephalic index* may be computed readily by multiplying the width by 100 and dividing by the length. This index is considered one of the most important of those used in anthropometry. By it the type of head may be determined as follows: if the index is less than 75, *S* is long-headed (dolichocephalic); if 75-80.9, *S* is 'medium' (mesocephalic); if 81-86.9, *S* is broad-headed (brachycephalic); if 87 or over, *S* is excessively broad-headed (hyperbrachycephalic).[1]

RESULTS. — (1) *Norms.* Typical head measurements are those made by Boas, West, Chamberlain and others upon Worcester school children, and reported by West: these are reproduced in Table 15.

For comprehensive summaries of the results of various investigators, consult Ernst.

(2) It was supposed by Galton that university students, whose heads were found to continue growth after the age of 19, presented an exception to the growth of skull of the general population, but it is more probable that in all males length of head continues to increase until the age of 21; in girls, maximal length of head is practically attained at 18. The growth, both of length and width of head, is very irregular, *i. e.*, periods of growth alternate with periods of cessation of growth.

---

[1]These points of divisions are not followed precisely by all investigators. Thus, Ernst counted as hyperdolichocephalic, -69.9; as dolichocephalic, 70-76.4; as mesocephalic, 76.5-80.9; as brachycephalic, 81.0-85.9; as hyperbrachycephalic, 86.0 +.

(3) *Dependence on sex.* Boys' heads are longer and wider than those of girls throughout the whole period of growth, and consequently throughout life. The heads of males are not only absolutely larger, but also relatively larger in comparison with their stature.

TABLE 15

Diameters of the Skull, in mm., and the Cephalic Index (West)

Age	Average Length		Average Width		Cephalic Index	
	Boys	Girls	Boys	Girls	Boys	Girls
5	176	174	140	138	79.56	79.40
6	177	172	142	139	78.94	79.60
7	179	175	142	140	79.42	80.02
8	180	174	143	141	78.71	80.41
9	181	176	144	140	79.63	79.71
10	182	177	145	142	80.30	79.46
11	183	180	144	142	78.80	78.90
12	183	180	145	143	79.40	79.40
13	184	181	147	145	79.50	79.60
14	187	183	147	144	78.60	79.00
15	188	184	148	146	78.59	78.99
16	191	184	149	144	77.81	78.48
17	189	185	150	146	78.34	78.50
18	192	186	151	147	78.88	79.36
19	192	183	150	145	78.33	79.68
20	195	182	152	147	77.88	79.41
21	192	186	153	145	79.29	78.36

(4) *Dependence on height.* In Oxford students Schuster found a correlation between stature and head length of 0.31, between stature and head width of 0.14. Analogous correlations reported by Pearson for Cambridge, Eng., students are 0.28 and 0.15, respectively, and for 3000 criminals 0.34 and 0.18, respectively.

(5) *Dependence on intelligence.*[1] With these measurements chief interest attaches to the question of correlation between skull size and shape and intellectual ability. The oft-quoted work of Galton comprises an investigation (assisted by Venn) of Cambridge students, who were divided on the basis

[1]Compare on this point the results listed in Test 4.

of intellectual ability into three groups — (a) "high-honor men," (b) "remaining honor men," (c) "poll" or "pass" men. When length, breadth and height of skull were multiplied together, Galton found that at 19 years of age the high-honor men had heads considerably larger than others, and that this difference was reduced by about one-half at the age of 25, and he concluded that high-honor men are presumably, as a class, both more precocious and more gifted throughout than others.

Vaschide and Pelletier measured 400 pupils of both sexes in the elementary schools of Paris, and, upon comparison with their teachers' estimates of ability, concluded that pupils of superior intelligence have on the average larger heads than the less intelligent, that this difference is independent of general physical development, and that height of head is the most important single cranial symptom of ability.

This last conclusion, however, does not accord with that reached by Binet in his earlier work (1900). He reports that the head of the unintelligent is smaller than that of the intelligent child in all dimensions save in vertical diameter and distance from the base of the nose to the end of the chin, though the differences are but slight and somewhat uncertain. If, however, exceptionally bright children (*enfants d'élite*) are compared with exceptionally dull children (*enfants arriérés*), differences averaging 3·4 mm. or more appear, particularly in transverse dimensions, *i. e.*, in breadth. Exceptionally bright children distinctly surpass average children, but the latter do not differ so much from dull children. In brief, then, exceptionally bright children are characterized by unusually wide heads.

The most elaborate statistical investigation by the group method is that of Karl Pearson, who collected data from 5000 school children in addition to the 1000 Cambridge undergraduates, and who classified these 6000 cases into four (or six) mental groups. From his data (Table 16) it is clear that, while there exists a sensible relationship between skull dimensions and intellectual ability, it is very small and not sufficient for any practical predictions: indeed, he discovers that the relation between color of hair and intelligence is just as good as that

between size and shape of skull and intelligence (14, p. 128). Miss Lee could find no correlation between estimated cranial capacity and estimated intellectual capacity of 60 men and 30 women.

TABLE 16

Correlation Ratio Between Intelligence and Size and Shape of Head (Pearson)

INTELLIGENCE AND	CAMBRIDGE GRADUATES			SCHOOL BOYS AGED 12 YEARS			SCHOOL GIRLS AGED 12 YEARS		
	NO.	RATIO	P. E.	NO.	RATIO	P. E.	NO.	RATIO	P. E.
Cephalic Index	1011	—.061	.02	2345	—.041	.01	2226	.067	.01
Length of Head	1011	.111	.02	2298	.139	.01	2188	.084	.01
Breadth of Head	1011	.097	.02	2299	.109	.01	2165	.113	.01

On the other hand, Porter found in width of head a positive correlation with school grade. His method of arraying his data to show this principle is illustrated in Table 17, in which all the girls aged 12 and all the boys aged 10 are distributed according to their school grades. It is then seen that those children of a given age in an advanced school grade have, on the average, broader heads than those in a lower grade.

TABLE 17

Breadth of Head by School Grade (Porter)

SCHOOL GRADE	BOYS AGED 10		GIRLS AGED 12	
	CASES	AVERAGE	CASES	AVERAGE
		mm.		*mm.*
I	92	145.86	—	—
II	408	146.73	68	143.68
III	397	146.48	193	144.77
IV	170	147.21	343	144.94
V	—	—	217	145.50
VI	—	—	89	147.64

Similarly, according to MacDonald, "dolichocephaly increases in children as ability decreases. A high percentage of dolichocephaly is, to a certain extent, a concomitant of mental dullness." "Unruly boys have a large percentage of long-headedness."

(6) The extensive study of *backward and feeble-minded children* conducted by Simon confirms in a general way Binet's results with normal children of varying intelligence, and also shows that the abnormal child only rarely has a head that is average in size and shape. In other words, variability is a chief trait in such heads.

Binet (3), in 1904, sought to establish certain limiting boundaries (*frontières anthropométriques*) for abnormal children. Figures were assigned for stature, length of head, breadth of head and for the sum and the differences of these two diameters, to apply to boys at each age from 6 to 18—a scale which reminds one of his well-known metric scale of intelligence for various ages as based upon mental tests. Trial by Decroly convinced him that these boundaries were of real practical value in raising a presumption of mental inferiority.

Still later, in 1910, Binet (4) contrived yet another series of limits by adding together five cranial measures, from the use of which he concluded that children taken in groups designated as bright, medium, dull and backward will show characteristic distributions in terms of this combination index. For any individual child, however, he would predict mental inferiority only if the child's index showed a "cephalic retardation" of 6 years or more in the scale of norms, but a retardation of from 3 to 6 years, when conjoined with serious school retardation or unfavorable outcome in psychological examination, would possess a certain confirmatory value. In general, then, Binet concludes that size of head only roughly indicates size of brain, and that size of brain is but one factor determinative of intelligence—quality of brain is, of course, the other.

(7) *Cephalic index.* Norms for the cephalic index of American children are presented in Table 15. The measurements, by Engelsperger and Ziegler, of 238 boys and 238 girls of the entering classes (average age 6 years, 4.5 months) in the schools of Munich, furnish results that deviate somewhat from those just cited for American children, as is seen clearly by a comparison of Tables 15 and 18.

It is of interest to note that no cases of dolichocephaly were found, but that these children were decidedly brachycephalic.

(8) *Dependence of cephalic index on race and environment.* The general conclusion that the cephalic index tended to assume typical proportions with each race has been recently shaken by the work of Boas, who measured the heads of American immigrants and their children born in this country. Boas declares that "the head form, which has always been considered as one of the most stable and permanent characteristics of hu-

TABLE 18

Skull Dimensions and Proportions of Entering Classes at Munich (Engelsperger and Ziegler)

	LENGTH			BREADTH			INDEX		
	Mean	Max	Min.	Mean	Max.	Min.	Meso.	Brac'y	Hyp'b
Boys	170.35	185	148	146.34	160	133	6.3	54.6	39.1
Girls	165.83	186	151	142.97	159	130	6.7	45.8	47.5

man races, undergoes far-reaching changes, due to the transfer of the races of Europe to American soil. The East European Hebrew, who has a very round head, becomes long-headed; the South Italian, who in Italy has an exceedingly long head, becomes more short-headed; so that both approach a uniform type in this country." These changes "develop in early childhood and persist throughout life," and the influence of American life "increases with the time that the immigrants have lived in this country before the birth of their children." But Radosavljevich criticizes severely Boas' methods of collecting and treating these data, and declares his results meaningless.

NOTES.—Heads of unusual shape or size, irregular or deformed, should receive especial care in measurement, and a descriptive note should be appended to the record.

Attempts to record the shape and size of the skull by means of the registering 'conformateur' used by hatters have usually been relinquished, because the hair interferes too much with exact determination. This instrument might, however, be of service in preserving a rough 'picture' of heads of unusual size or proportions.

Since brain size and form are only roughly indicated by the exterior dimensions of the head, while intelligence is conditioned primarily by the elaborateness of the finer nerve structure and not (save in pathological cases of hypertrophy or developmental arrest) by the gross size or form of the brain, it is scarcely necessary to call attention to the absurdity, *a fortiori*, of the claims of phrenology.

REFERENCES

(1) A. Bertillon, Signaletic instructions; including the theory and practice of anthropometrical identification. Eng. tr., Chicago and New York, 1896. Pp. 260.

(2) A. Binet, Recherches sur la technique de la mensuration de la tête vivante, and four other articles on cephalometery. *AnPs*, 7: 1900 (1901), 314-429.

(3) A. Binet, Les frontières anthropométriques des anormaux. *BuSocEtPsEnf*, 1904, p. 430.

(4) A. Binet, Les signes physiques de l'intelligence chez les enfants. *AnPs*, 16: 1910, 1-30, especially 3-12.

(5) F. Boas, Changes in bodily form of descendants of immigrants. Washington, 1910. (Immigration Commission, Senate Doc., No. 208). (Also in his book, The mind of primitive man, N. Y., 1911).

(6) O. Decroly, Les frontières anthropométriques des anormaux d'apres M. Binet appliquées a des enfants arriérés de Bruxelles. Extrait des Annales publiés par la *Société roy. des sci. méd. et nat. de Bruxelles*, 14: 1905. Pp. 35.

(7) A. Engelsperger and O. Ziegler, Beiträge zur Kenntniss der physischen und psychischen Natur des sechsjährigen in die Schule eintretenden Kindes. Anthropometrisches Teil. *EPd*, 1: 1905, 173-235.

(8) Lucy Hoesch Ernst, Das Schulkind in seiner körperlichen u. geistigen Entwicklung. I Teil. Leipzig, 1906, pp. 100-122.

(9) F. Galton, Head growth in students at the University of Cambridge. *Nature*, 38: 1888, 14-15; 40: 1889, 317-8. Also Cambridge anthropometry, *Nature*, 41: 1890, 454.

(10) ——— Historique des recherches sur les rapports de l'intelligence avec la grandeur et la forme de la tête. *AnPs*, 5: 1898 (1899), 245-298.

(11) Alice Lee, Study of the correlation of the human skull. *Sci.*, n. s. 12: 1900, 946-9.

(12) A. MacDonald, Experimental study of children, etc. Reprint of chs. 21 and 25 of *RepComEd*, 1899.

(13) E. Meumann, Intelligenzprüfungen an Kindern der Volksschule. *EPd*, 1: 1905, 35-100, especially 58 ff.

(14) K. Pearson, On the relationship of intelligence to size and shape of head, and to other physical and mental characters. *Biometrika*, 5: 1906, 105-146.

(15) W. T. Porter, The growth of St. Louis children. *Trans. Acad. Sci. St. Louis*, 6: 1894, 263-426.

(16) W. T. Porter, Physical basis of precocity and dullness. *Ibid.*, 161-181.

(17) P. R. Radosavljevich, Professor Boas' new theory of the form of the head.—A critical contribution to school anthropology. *AmAnt*, n. s. 13: 1911, 394-436 (bibliography of 80 titles).

(18) E. Schuster, First results from the Oxford anthropometric laboratory. *Biometrika*, 8: 1911, 40-51.

(19) Simon, Recherches cephalometriques sur les enfants arriérés de la colonie de Vaucluse. *AnPs*, 7: 1900 (1901), 430-489.

(19) N. Vaschide et Pelletier, Recherches sur les signes physiques de l'intelligence. *RPh*, 1903-4.

(20) G. West, Worcester school children; the growth of the body, head and face. *Sci.*, 21: 1893, 2-4.

TEST 4

Girth of the skull.—This measurement is less in favor with investigators than those just described, because of the variable factor of the hair, mentioned in the preceding test.

INSTRUMENT.—Anthropometric measuring tape (Fig. 10).

METHOD.—*E* stands at the right of *S*, who is seated. *E* holds

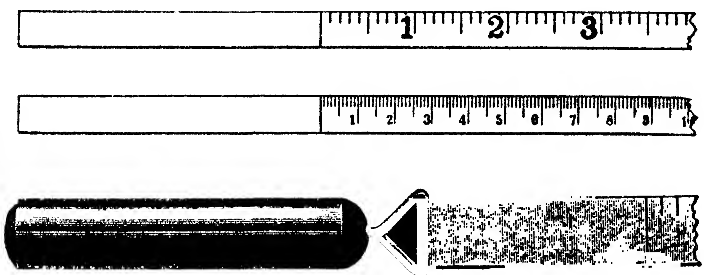


FIG. 10. ANTHROPOMETRIC TAPE.

the tape with the thumb and forefinger of each hand at a length approximately that of the distance to be measured. He then lifts the tape over *S*'s head, keeping it horizontal, and applies it about the head at such a height as to pass around the largest part—over the frontal prominences and over the occipital prominences. The tension of the tape is regulated by observation of the spring-indicator.

RESULTS.—(1) *Norms.* Measurements by MacDonald of the circumference of the heads of 7953 boys and 8520 girls in Washington, D. C. supply the norms (converted into mm. from his original tables in inches) embodied in Table 19. Comparison with the results published by Ernst that have been obtained by Quetelet, Landsberger, Hrdlicka and herself will show that MacDonald's norms are somewhat larger than those obtained by these investigators. For ages above 18 we may refer to the results of Schuster, who found that the mean circumference of head in Cambridge undergraduates during the ages 18 to 23 and over ranged roughly from 560.76 to 565.53 mm., with a general average of 563.10 mm.

TABLE 19.

Circumference of the Head, in mm. (MacDonald)

AGE	BOYS	GIRLS	AGE	BOYS	GIRLS
6	513.6	506.2	13	533.7	532.1
7	519.4	506.5	14	538.7	538.0
8	520.9	511.6	15	544.8	540.5
9	523.5	515.4	16	550.4	543.0
10	526.5	518.9	17	555.5	547.4
11	528.8	521.7	18	556.5	548.6
12	531.9	527.8			

(2) *Dependence on sex.*[1] The head circumference of males is larger than that of females at all ages, both absolutely and relatively (in comparison with stature). MacDonald, however, found an exception in the case of colored children (see his Diagrams 2 and 3, pp. 1017-1018). Some investigators, notably Möbius, have argued from these facts to a general mental superiority of males over females.

(3) *Dependence on social status.* Children of the non-laboring classes have a larger head circumference than children of the laboring classes (MacDonald, Diagram 1, p. 1016).

(4) *Dependence on stature.* If the head circumference be multiplied by 100 and divided by the stature, the resultant ratio will be found to diminish as stature increases. Thus for 9-year-old boys with statures of 121 to 122 cm. the ratio is 42.3 to 42.9; for 10-year-old boys with statures of 124.8 to 127.3 cm. the ratio is 41.3 to 41.7, while with 11-year-old boys with statures 130 to 135.9 cm. the ratio is 39.0 to 39.9 (Ernst, from the data of different investigators).

(5) *Dependence on intelligence.* Möbius, as intimated above, believes that, at least in the case of normal adults, mental capacity tends to correlate with skull circumference. MacDonald, who related head girth with mental ability estimated by the teacher, concludes that, as the girth increases, mental ability increases, provided that one and the same race be under

[1]Consult MacDonald (pp. 1016ff.) for an extended discussion of the relation of circumference of head to sex, nativity, race, sociological condition and mental ability.

consideration. Dr. Ernst, however, raises the question whether this apparent correlation is not primarily determined by the relation above mentioned between size of head and social milieu.

Of special interest are the several articles by Bayerthal. In 1906, he measured the skull circumference of 234 boys and 153 girls (ages 7.5 to 8.5 years) and related these measures with school standing by classifying both sexes into five groups, as "very good" (I), "good" (II), "good on the whole" (III), "satisfactory" (IV), and "more or less unsatisfactory" (V). The results tend to confirm the existence of a positive correlation between skull circumference and general ability. Thus, the average skull circumferences were, for boys, 51.46, 50.93, 50.33, 49.60 and 49.60 cm., and for girls, 50.00, 49.83, 49.44, 49.16 and 48.84 cm., for the groups I to V, respectively.

In 1910 Bayerthal reaches the general conclusion that large heads may have all grades of intelligence from genius to idiocy, medium-sized heads may have intelligence above the average, though highly-developed intelligence is rare, while very small heads are never coupled with high-grade intelligence. There is, in other words, a one-way correspondence: very small heads exclude high intelligence, but very large heads do not guarantee it. To define more exactly what Bayerthal means by small heads we may reproduce his conclusions in tabular form as follows:

Very good intellectual capacity appears to be

	POSITIVELY EXCLUDED			ONLY RARELY PRESENT		
If at ages	7	10	12-14	7	10	11-12
In boys the girth is less than	48	49.5	50.5	50	52	52.5
In girls the girth is less than	47	48.5	49.5	49	51	51.5

If these boundaries (comparable to the anthropometric frontiers proposed by Binet, Test 3) are confirmed by further investigation, we may realize Bayerthal's expectation that the school physicians of the future may, at the end of the first school year, pick out by skull measurements some at least of

those who are intellectually incapable of carrying the work of the public school.

(6) Children with abnormalities are inferior in head circumference to normal children.

(7) In an examination of 60 juvenile delinquents, Dawson found the average circumference of head less than that of normal children of the same age: in 64 per cent. of the cases studied the circumference was from 1.7 to 5.2 cm. less than the mean for normal children.

REFERENCES

(1) Bayerthal, Kopfumfang und Intelligenz im Kindesalter. *EPd*, 2: 1906, 247-251.

(2) Bayerthal, Weitere Untersuchungen über Kopfumfang und Intelligenz im Kindesalter. *EPd*, 3: 1906, 238-242.

(3) Bayerthal, Weitere Untersuchungen über die Beziehungen zwischen Schädelumfang und Intelligenz im schulpflichtigen Alter, *ZEPd*, 5: 1907, 223-230.

(4) Bayerthal, Kopfgrösse und Intelligenz im schulpflichtigen Alter. *ZEPd*, 10: 1910, 197-218.

(5) G. E. Dawson, A study in youthful degeneracy. *PdSe*, 4: 1896, 221-258.

(6) Lucy Hoesch Ernst, Das Schulkind in seiner körperlichen u. geistigen Entwicklung. I Teil. Leipzig, 1906, pp. 92-99.

(7) A. MacDonald, Experimental study of school children, etc., reprint of chs. 21 and 25 of *RepComEd*, 1899.

(8) Möbius, Geschlecht und Kopfgrösse. Halle, 1903.

(9) E. Schuster, First results from the Oxford anthropometric laboratory. *Biometrika*, 8: 1911, 40-51.

CHAPTER V.

TESTS OF PHYSICAL AND MOTOR CAPACITY.

The title 'physical and motor capacity' is here used as a convenient and practical phrase to cover a number of tests which have often been classified under diverse rubrics, such as strength tests, motor tests, physical tests, tests of physiological condition, etc. All of the tests here described differ from the anthropometric tests of Chapter IV in that they measure not mere size or dimension, but functional, especially muscular capacity. They differ from the tests of Chapter VI, many of which might equally well be said to measure physiological condition or capacity, *e. g.*, the test of visual acuity, in that they are primarily tests of motor rather than of sensory capacity.

The first test described, that of vital capacity (often loosely termed lung capacity), is, perhaps, not so obviously a test of muscular efficiency as are the four strength tests that follow. It is, however, clearly a test of physical capacity dependent upon movement. The tests of quickness, accuracy and steadiness of movement are frequently placed in a class by themselves under the rubric 'motor tests,' but they are easily subsumed under the title here employed.

Reaction-time would by many be considered a test of quickness of movement; but it is so largely dependent upon complex psychological conditions, particularly upon the instructions, the direction of attention and the type of stimulus employed, that it belongs rather to the experimental examination of action than to the measurement of physical capacity as such.[1]

These tests of physical and motor capacity have become prominent chiefly because of their employment in the study of the correlation of physical and mental ability. For this pur-

[1]See an article by the writer, "Reaction-Time as a Test of Mental Ability." *AmJPs*, 15: 1904, 489.

pose they are commonly used in conjunction with the anthropometric tests already described and with various tests of general intelligence or mental ability to be described later.

These tests have also an obvious and direct application in the study of various problems of hygiene, physical culture, etc., while some of them bid fair to serve a useful purpose in the selection of applicants for certain industrial vocations.

TEST 5

Vital capacity.—Vital capacity, also termed breathing capacity and differential capacity, is the maximal volume of air that can be expired after taking a maximal inspiration. It is not identical with lung capacity, because a certain amount of air, termed the residual air, always remains in the lungs.

Vital capacity is considered an important index of general physical condition and capacity, and has, accordingly, found a place in nearly all measurements of school children in which the physical status has been examined. It is affected by sex, age, stature, posture, occupation, amount of daily physical activity and by disease, and may be markedly increased (*e. g.*, 300 cc. in three months) by various forms of physical exercise which demand active respiration.

The ratio of vital capacity to weight is termed the *vital index,* and is held to be of extreme significance, because it expresses the balance between bodily size and the rate and completeness with which oxidization of the blood is, or may be, effected. A high vital index is undoubtedly a preventive of auto-intoxication, gives increased resistance to disease, and is the root of endurance under effort. Thus, athletic training consists primarily in the reduction of weight and the increase of breathing capacity.

APPARATUS.—Spirometer, preferably of the wet type (Fig. 11), fitted with detachable wooden mouthpiece. Extra mouthpieces.

METHOD.—See that *S*'s clothing is perfectly loose about his neck and chest. Instruct him to stand upright, to take as full an inspiration as possible, and then to blow all the air he can

not too rapidly, into the spirometer. Also caution him to take care that no air escapes about the mouthpiece.

Two or three trials may be allowed, and the best record set down.

After *S*'s record is made, discard the mouthpiece and insert a new one into the rubber tube.

RESULTS.—(1) *Norms.* The norms of vital capacity presented in Table 20 are those established by Smedley with Chicago school children. The distribution of Smedley's data by

FIG 11. WET SPIROMETER.
Graduated in cubic inches and cubic decimeters.

percentile grades is shown in Figs. 12 and 13. For additional results Ernst's monograph may be consulted for vital capacity and chest measurements of Zurich children and summarized tables of the work of Pagliani, Kotelmann and Gilbert, while in Hastings (pp. 79-112) will be found an excellent set of anthropometric tables for vital capacity and various other phys-

ical measurements for boys and girls of each age from 5 to 20, arranged for 8 characteristic heights at each age. The average capacity reported by Vierordt for adults—3400 cc. for men and 2400 cc. for women—is noticeably lower than the Chicago norms for boys in late adolescence, and much lower than the average capacity, 4315 cc., found by Schuster in undergraduates at Cambridge University.

TABLE 20

Norms of Vital Capacity, in Cubic Centimeters (Smedley)

AGE	BOYS	GIRLS	AGE	BOYS	GIRLS
6	1023	950	13	2108	1827
7	1168	1061	14	2395	2014
8	1316	1165	15	2697	2168
9	1469	1286	16	3120	2266
10	1603	1409	17	3483	2319
11	1732	1526	18	3655	2343
12	1883	1664			

(2) *Dependence on sex.* All investigators agree that boys have a larger vital capacity than girls at all ages, and that men, similarly, have a larger capacity than women. Even if we compare men and women of the same height, the former surpass the latter by about the ratio 10 : 7.5. The increase of vital capacity in boys is slow and steady during the years 6 to 12, but very marked during the next four years, whereas the most rapid increase in girls is during the years 11 to 14. In both sexes, then, these periods of rapid increase coincide with the periods of rapid growth in height and weight.

(3) *Dependence on height.* The norm is conditioned by height. For each centimeter of increase or decrease of stature above or below the mean there is a corresponding rise or fall of the vital capacity, amounting in men to 60 cc., in women to 40 cc. This correlation with height varies somewhat at different ages. Thus, according to Wintrich, the average vital capacity for each centimeter of height is, from 8 to 10 years, 10 cc., from 16 to 18 years, 20.65 cc., and at 50 years, 21 cc.

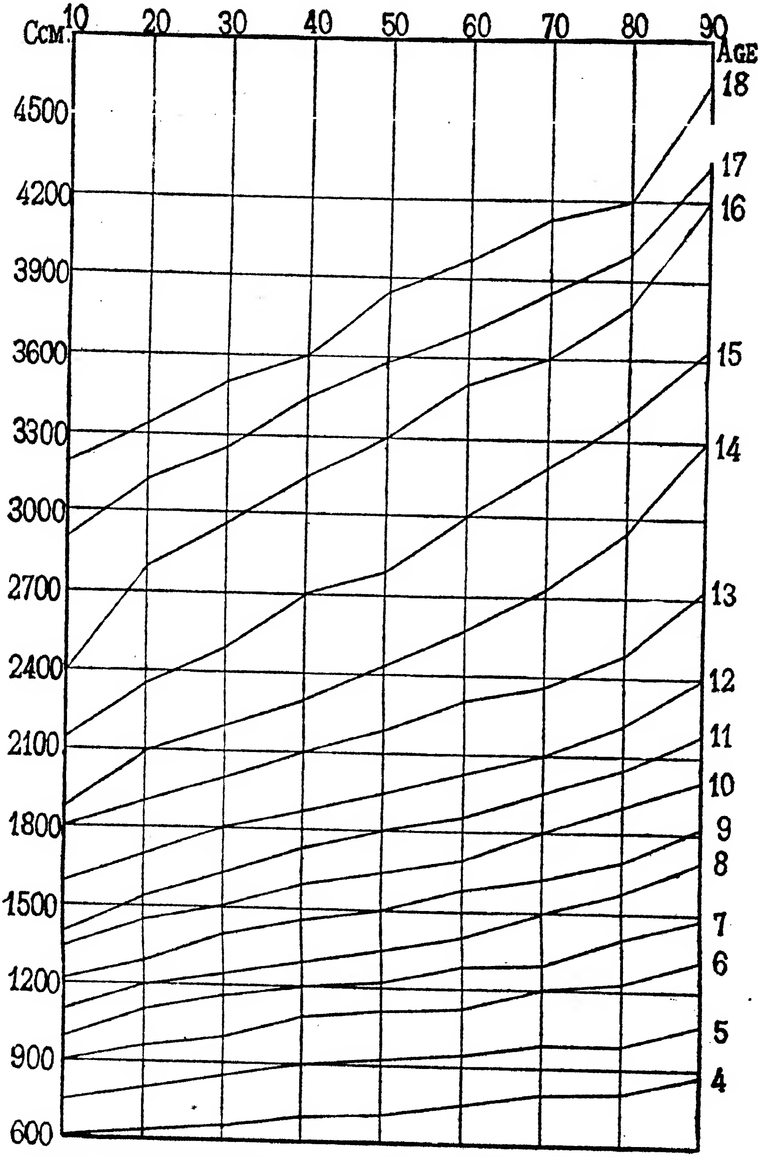


FIG. 12. PERCENTILES OF VITAL CAPACITY FOR BOYS (SMEDLEY)

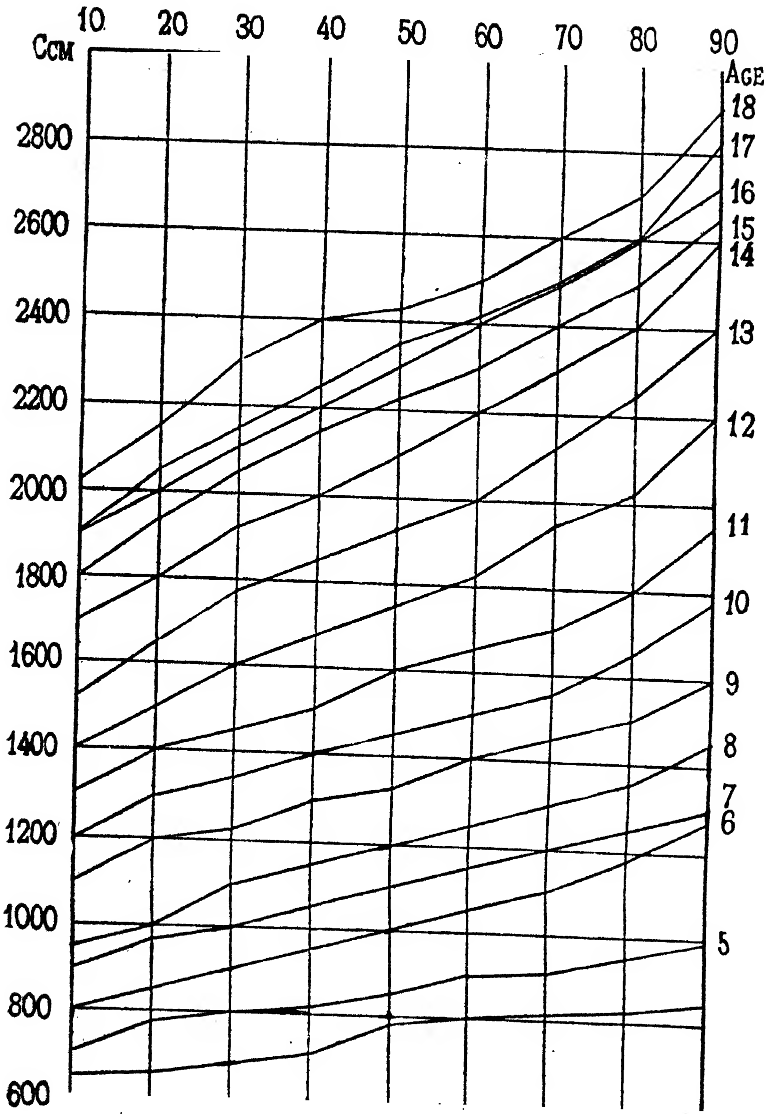


FIG. 13 PERCENTILES OF VITAL CAPACITY FOR GIRLS (SMEDLEY)

Schuster found the general correlation between vital capacity and height to be + .57, and that between the same capacity and weight to be + .59. It follows that in estimating the status of a given individual's vital capacity allowance must be made for his size of body as well as for his age.

(4) *Dependence on age.* The general dependence on age is indicated in the norms already given. It may be mentioned further that the most marked individual differences appear at the ages when the period of most rapid growth terminates (Smedley). Concerning the further development of vital capacity after the period of adolescence, Beyer, from his study of naval cadets, concludes that maximal capacity is reached at 19, but other authorities place the maximum at 35, with an annual decrease of about 32 cc. thereafter up to the age of 65.

(5) *Dependence on mode of life.* It has been pointed out above that vital capacity may be decidedly increased by forms of physical exercise that tend to induce forced respiration. Capacity is likewise proportionately reduced in men who live a sedentary life. It is also reduced by any circumstance which interferes with the free expansion of the thorax, such as tight clothing, tuberculosis of the lungs, visceral tumors, etc.

TABLE 21

Value of the Vital Index, when Weight is Taken as Unity (Kotelmann)

AGE	INDEX	AGE	INDEX	AGE	INDEX	AGE	INDEX
9	69.32	12	67.51	15	63.18	18	64.28
10	69.37	13	66.75	16	65.94	19	66.22
11	69.18	14	64.07	17	65.77	20	65.01

(6) The relation between weight and vital capacity, *i. e.*, *the vital index*, presented in Table 21, is that found by Kotelmann, also given by MacDonald. The ratio expresses the relation in terms of kg. of weight and cc. of vital capacity. It will be seen that the weight of the body normally increases with age somewhat faster than the vital capacity. If height be similarly treated, it will be found, on the contrary, that vital capacity increases with age faster than it increases.

(6) *Relation to mental ability.*[1] Gilbert found the correlation between vital capacity and mental ability indifferent or negative: during the years 10 to 15, indeed, duller children, he found, had the larger capacity

On the other hand, Smedley found a positive correlation between school standing and vital capacity, whether he took the distribution through the grades of all pupils of a given age, or computed the average school-grade of those who stood at various percentiles of vital capacity, or compared those at and above grade with those below grade at each age. Moreover, the same investigator found that pupils in the John Worthy School (incorrigibles, truants, etc.) were, from the age of ten up, inferior in vital capacity to children in the other schools, and that the inferiority became more noticeable with age. I am informed also by Goddard that the spirometer records of feeble-minded children are invariably below normal, and DeBusk's study of 105 Colorado pupils showed greater divergences in averages for vital capacity than for height and weight when accelerated, normal and retarded groups were compared. Almost without exception, he says, pedagogically retarded pupils show a vital capacity under the average. DeBusk also computed the vital index in terms of cc. per pound of body weight, and found the index of children pedagogically 'at grade' constant at about 25-26 cc. per pound for all ages. In the third grade, to take an instance, the general average for the index is 24.4 cc., but it falls to 23 cc. for pupils one year retarded and to 20 cc. for pupils two years retarded.

NOTES.—The dry spirometer is less expensive than the wet in first cost, and is more portable, but it has the disadvantage of getting out of repair easily. Its readings are apt to run slightly higher than those of the wet spirometer. Dr. Ernst believes that differences in apparatus are so serious that figures obtained with different types of spirometer are not directly comparable.

The mouthpiece of the ordinary spirometer forms an excellent medium for the dissemination of bacteria. For this reason

[1] See also the discussion of correlation with mental ability in Test 1.

the detachable mouthpieces are imperative if hygienic conditions are to be assured.

There is a certain knack in making a maximal spirometer record; some children may exhibit it; others not. In particular, to get a good record, the expiration must be neither too fast nor too slow, and an extra effort must be made just at the end of both inspiration and expiration to utilize the available lung capacity to the utmost.

REFERENCES

(1) H. G. Beyer, The growth of United States naval cadets. *Proc USNInst*, 21: No. 2, 1895.
(2) B. W. DeBusk, Height, weight, vital capacity and retardation. *PdSe*, 20: 1913, 89-92.
(3) Lucy Hoesch Ernst, Das Schulkind in seiner körperlichen und geistigen Entwicklung. I Teil. Leipzig, 1906. Pp. 67-9. (For diameters and girth of chest, see also pp. 59-67.)
(4) J. A. Gilbert, Researches on the mental and physical development of school children. *SdYalePsLab*, 2: 1894, 40-100.
(5) J. A. Gilbert, Researches upon school children and college students. *UnIowaSdPs*, 1: 1897, 1-39.
(6) W. W. Hastings, A manual for physical measurements. Springfield, Mass., 1902.
(7) L. Kotelmann, Die Körperverhältnisse der gelehrten Schüler des Johanneums in Hamburg. *Zeits. d. Königl. Preus. statist. Bureaus*, 1879.
(8) A. MacDonald, Experimental study of school children, etc. Reprint of chs, 21 and 25 of *RepComEd*, 1899.
(9) E. Schuster, First results from the Oxford anthropometric laboratory. *Biometrika*, 8: 1911, 40-51.
(10) F. Smedley, Rept. dept. child-study and pedagogic investigation. Reprint from 46th *An. Rep. Brd. Educ., Chicago*, 1899-1900. Also *RepComEd*, 1902, i., 1095-1115.
(11) F. Smedley, do., No. 3, 1900-1901. Also *RepComEd*, 1902, i., 1115-1138.

TEST 6

Strength of grip.—This test has been used to secure an index of general bodily strength, to secure an index of right-handedness[1] (in conjunction with Tests 10 to 12), and for comparative purposes generally. It may be modified to secure an index of endurance or fatigue (Test 9), or combined with other forms of strength measurement (Tests 7 and 8).

[1]The terminology of right and left-handedness is at present somewhat confused (Cf. E. Jones. *PsBu*, 6: April, 1909). The terms 'index of unidexterity' and 'index of dextrality' have been used by some writers as equivalent to 'index of right-handedness.' 'Dextrality' is here used to indicate the superiority of one hand (whether right or left) over the other.

APPARATUS. — Improved form of Smedley's dynamometer (Fig. 14). Millimeter rule.

METHOD. — With the millimeter rule, measure the distance from where *S*'s thumb joins his hand to the end of his fingers. Adjust the dynamometer by whirling the inner 'stirrup' until the scale on the outer stirrup indicates one-half this distance. This should bring the second phalanx to bear against the inner stirrup, and will ordinarily prove to be the optimal adjustment; if not, it may be modified to suit *S*'s inclinations. Then set the instrument by means of the clutch, so that the inner stirrup cannot twist while in use, and record the adjustment by reference to the scale upon the stirrup.

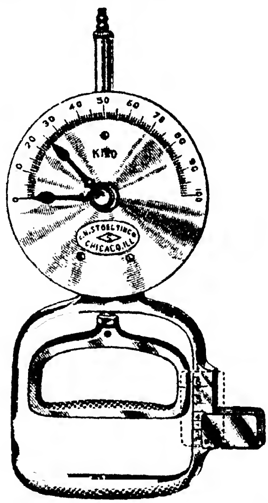


FIG. 14. DYNAMOMETER AND DYNAMOGRAPH, AFTER SMEDLEY, IMPROVED.

Illustrate the use of the instrument to *S*; especially make clear that the lower pointer will register the grip, so that he does not have to continue his effort while the scale is read.

Allow three trials with each hand, right and left alternately, but introduce a brief pause, say 10 seconds, between each trial to avoid excessive fatigue. Have *S* exert his maximal grip, and in each trial encourage him to do his best. Record the amount registered at each trial; but, for ordinary purposes, use in subsequent computation only the highest record for each hand.

Records obtained upon the dynamometer, as nearly all investigators have pointed out, are very liable to be affected by subjective factors. Individuals who are, by assumption, exerting maximal effort, can often shove the pointer a few kilograms farther over the scale if some stimulating appeal is made to them. In order to get this last ounce of energy into the record some experimenters, *e. g.*, Binet and Vaschide, Engelsperger and Ziegler, recommend that all dynamometric records be made in the presence of other *S*'s, so that competition shall induce the best possible efforts. For a similar purpose Wallin recommends that the Smedley instrument be held with the palm up and with the dial up, so that *S* may watch the movement of the pointer. Whatever be the conditions adopted by *E*, it is evident that the same conditions should prevail for all *S*'s.

Engelsperger and Ziegler also recommend that this test be repeated on several different days (at about the same hour each day) in order to avoid chance errors like temporary indisposition, etc.

Results.—(1) *Norms*. Tests of 2788 boys and 3471 girls at Chicago with the Smedley dynamometer yielded the norms of Table 22. The corresponding percentile distributions are shown in Figs. 15 and 16. For an extensive study of boys and girls aged 6 years, including tables of distribution, the work of Engelsperger and Ziegler should be consulted, while valuable tables, showing averages and standards for the 25th and the 75th percentile for boys and girls of each age from 5 to 20, as related to different statures, will be found in Hasting's *Manual* (pp. 79-112).

TABLE 22.

Norms of Strength of Grip, in kg. (Smedley)

AGE	BOYS		GIRLS	
	Rt. Hand	Lt. Hand	Rt. Hand	Lt. Hand
6	9.21	8.48	8.36	7.74
7	10.74	10.11	9.88	9.24
8	12.41	11.67	11.16	10.48
9	14.34	13.47	12.77	11.97
10	16.52	15.59	14.65	13.72
11	18.85	17.72	16.54	15.52
12	21.24	19.71	18.92	17.78
13	24.44	22.51	21.84	20.39
14	28.42	26.22	24.79	22.92
15	33.39	30.88	27.00	24.92
16	39.37	36.39	28.70	26.56
17	44.74	40.96	29.56	27.43
18	49.28	45.01	29.75	27.66

(2) *Dependence on age*. The general increase of strength with age is indicated in Table 22. It may be added that the

individual variation, like that for most physical traits, is more pronounced in early adolescence than at any other time.

(3) *Dependence on sex.* Boys are uniformly stronger than girls, and men than women. The difference is greater than the mere difference in bodily size, and is to be attributed partly to lesser practise, partly to intrinsically weaker muscles, and perhaps partly to the mental factor, *i. e.,* less ability or inclination to maximal exertion of actual muscular strength (Engelsperger and Ziegler). The divergence between the sexes becomes more pronounced at puberty when sex differences in general are accentuated.

(4) *Dependence on race.* According to Hrdlicka, colored children have a stronger grip than white children at all ages.

(5) *Dependence on season.* Schuyten (quoted by Engelsperger and Ziegler) took dynamometer records at several times during a year, and claims that there are seasonal variations, that muscular strength decreases from January to March, increases from April to June, probably decreases again from July to September, then increases from October to December.

(6) *Dependence on incentives.* If the test is taken under stimulating conditions, such as competition, personal encouragement, public announcement of records, etc., Binet and Vaschide found that the average grip was increased about 3 kg., or so much that the left hand surpassed the previous record of the right hand made without such incitement. Similarly, Schuyten (13) found that ennui, or loss of interest in successive tests, is sufficient to obscure the fatigue effect of a school session.

(7) *Dependence on social status.* Schuyten concludes that the children of well-to-do parents are stronger than the children of poor parents, and Hrdlicka found asylum children below average on account of their poorer circumstances and unfavorable environment in early childhood. On the other hand, the results of Engelsperger and Ziegler with 6-year-old Zurich children, and more especially of MacDonald with Washington children, suggest a contrary relation, which these authors attribute to the manual work done by poor children in partial self-support.

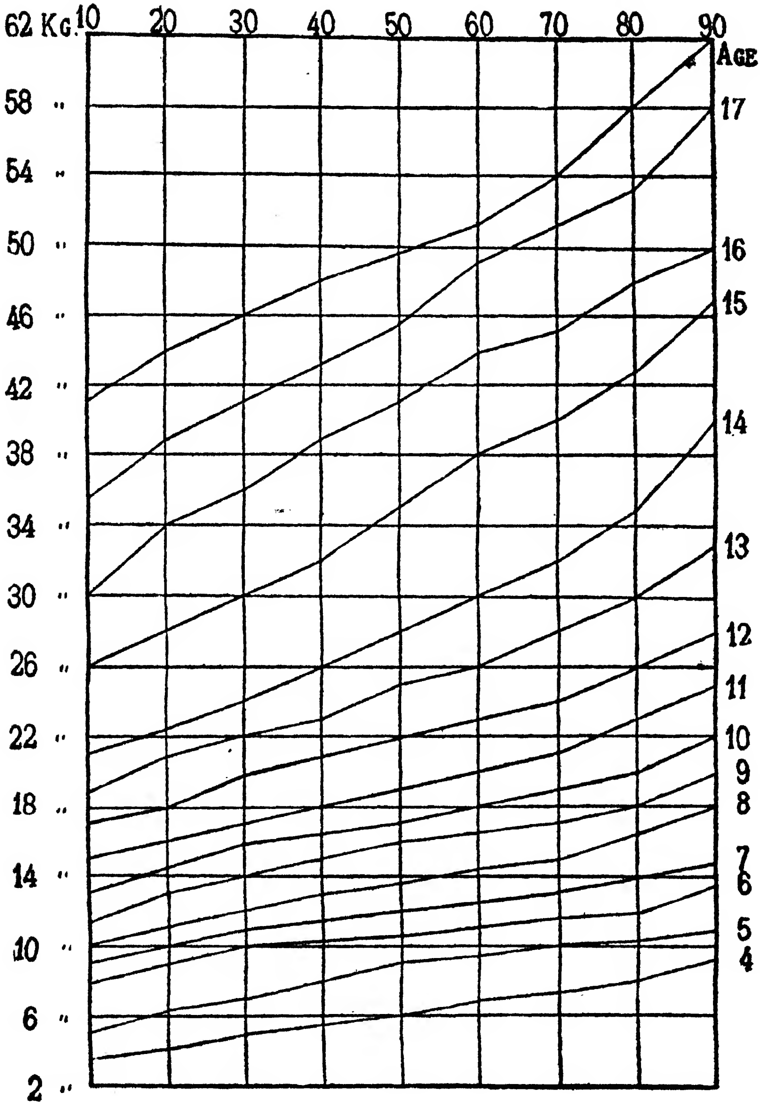


FIG. 15. PERCENTILES OF RIGHT-HAND GRIP FOR BOYS (SMEDLEY)

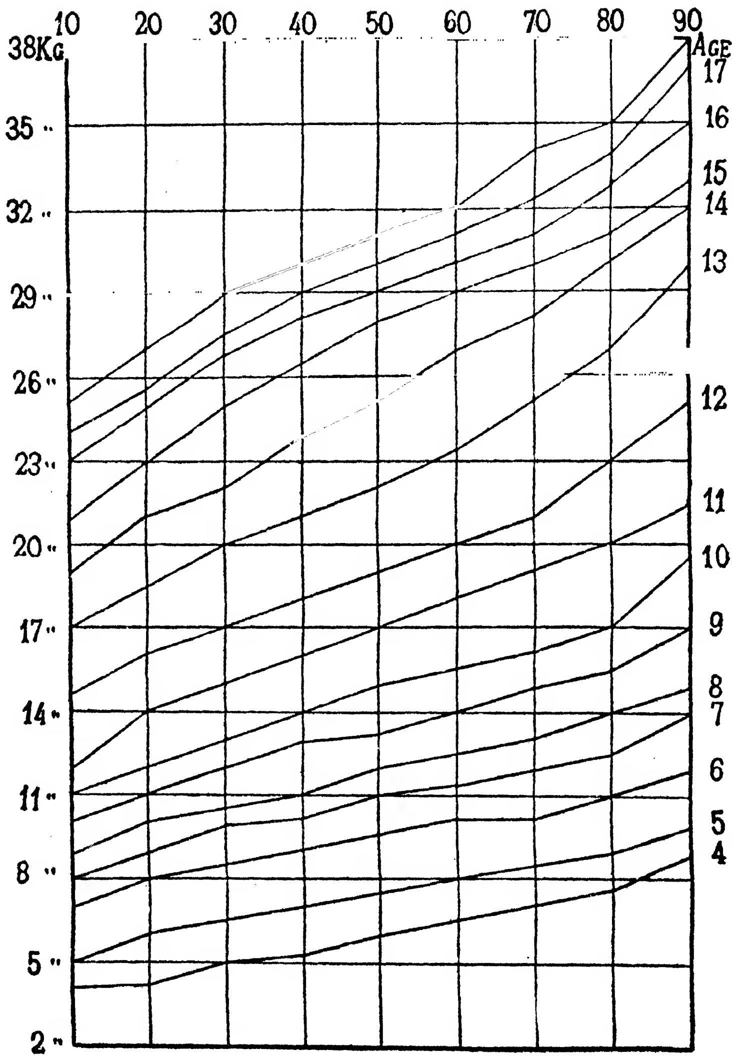


FIG. 16. PERCENTILES OF RIGHT-HAND GRIP FOR GIRLS (SMEDLEY)

(8) *Relation to intelligence.* MacDonald found no correlation between strength of hand and mental ability: the indications that, on the contrary, dull children tended to surpass average and bright children in this capacity are interpreted by him as due to the presence in the dull groups of numerous children from the poorer classes of the community, who tend, as just stated above, to be stronger in grip on account of manual activity.

Smedley, Schuyten and Miss Carman, however, have found evidences of positive correlation. The last named, from measurements of the grip of 1507 boys and girls aged 10 to 19 years, found that bright children exceeded dull children by an average of 3 kg. with the right, and 1 kg. with the left hand. In Chicago (14, 15) the existence of a positive correspondence between strength of grip and class standing was shown by three different methods, viz.: by the distribution of 12-year-old pupils by grades, by comparing the grip of those at and above grade with the grip of those below grade at each age, and by computing the average number of school grades that had been made by the various percentile groups (in strength) after sex and age had been eliminated. Schuyten, who estimated intelligence by school grade in relation to age, also found that those who are most intelligent are strongest.

(9) *Abnormal children.* Some confirmation for the positive correlation with mental status may, perhaps, be afforded by the statements of those who have worked with subnormal and abnormal children. Thus, Barr goes so far as to say (p. 162) that hand grasp and mental grasp go hand in hand, so that a test of grip is more serviceable than a test of language to diagnose mental status. Wallin's work with epileptics shows clearly that high-grade patients are stronger in grip than low-grade patients. Thus, those at the Skillman institution rated as morons surpass those rated as imbeciles by 11.6 kg. with the right and 12.2 kg. with the left hand; there is likewise a general increase of grip with increase in rated mental ages (based on the Binet-Simon tests). Again, Dawson found that juvenile delinquents have a mean strength of grip slightly less than normal children, and that 56 per cent. of them are

inferior to the normal by from 1.32 to 11.82 kg. Similarly, boys in the school for incorrigibles and truants at Chicago are, at every age from 9 to 17 and with either hand, less strong than normal boys, and this discrepancy increases very decidedly with age, *e. g.*, from 96.8 per cent. of the norm at the age of 9 to 63.2 per cent. of the norm at the age of 17.

(10) The *index of right-handedness, i. e.*, the percentage of strength of the left hand compared with the right, will be found to range, for any ordinary group of school children, between 91 and 96 per cent. However, occasional right-handed children may have an index of strength exceeding 1.00, *i. e.*, grip may be stronger with their left hand. According to Hrdlicka, this tendency to have an index contrary to expectation is peculiarly evident in left-handed persons, so that he estimates that nearly one-half of *bona-fide* left-handed persons may have a stronger grip with the right hand. Dr. Ernst found 3.7 per cent. left-handed boys and 4.8 per cent. left-handed girls, and these proved not always stronger with the left hand.

It is often asserted that degenerates tend to be left-handed. Wallin's averages for epileptics show a net preponderance of 0.5 kg. for the general average in favor of the left hand, though by no means all his *S*'s had an index above 1.00.

(11) *Dextrality, i. e.*, superiority of one hand over the other, is evident when the child enters school, but becomes increasingly evident as maturity approaches, and especially at puberty, so that a heightened difference in the strength of the hands may be regarded as one of the characteristic indications of pubertal change.

(12) *Relation between dextrality and intelligence.* The results just cited for epileptics might be restated by saying that in strength these children tend to be ambidextrous. This agrees well with the result found by Smedley that dull pupils are more nearly ambidextrous than average, and average than bright pupils. And, again, the boys in the John Worthy School are still more nearly ambidextrous than the dull pupils of the regular Chicago schools.

(13) *Relation between dextrality and absolute strength.* Just as mentally feeble children have less dextrality, so phys-

ically feeble children have less dextrality, so that these children may be said, as it were, to have "two left hands" (Binet and Vaschide).

(14) The exertion of maximal strength is commonly accompanied by characteristic poses, attitudes, facial contortions, grimaces, etc., which are, in general, evidences of the escape of uncontrolled energy through various motor paths. There appears to be an inverse relation between the strength and efficiency of the subject and the number and extent of these waste movements; these are correspondingly more evident when the muscles tire and *S* is unable to accomplish what he is attempting. In particular, a sort of foolish laugh is characteristic of this muscular inefficiency.[1]

NOTES.—The chief objections which have been made to the employment of the dynamometer are (1) that it is painful, particularly if a series of grips is taken; (2) that some *S*'s suffer from sweating of the hands, especially when excited, and that this causes the instrument to slip in their grasp; (3) that a wrong manner of holding the instrument may reduce the record, *e. g.*, by as much as 10 kg.; (4) that, owing to the large number of muscles concerned, a lack of proper coördination in their contraction may lower the record.

The painfulness of the dynamometer can be largely eliminated by proper construction; the Smedley instrument is much better than the Collin elliptical form so commonly used heretofore, unless the latter be reconstructed, as was done to advantage for the tests of Engelsperger and Ziegler. Moreover, if an extended investigation is to be undertaken, inurement to the pressure is rapidly developed (Bolton and Miller).

For the ascertainment of strength of grip, excessive perspiration can be avoided by simply drying the hands with a towel whenever necessary.[2]

The proper holding of the instrument is also largely depend-

[1]For a description, with photographic reproductions of these motor automatisms of effort, consult Binet and Vaschide.

[2]The author has found that slipping may be obviated by winding the two grip bars with ordinary bicycle tire tape. If the fresh tape is too sticky, apply a little talcum or powdered chalk after the winding is finished.

ent on proper construction, and in this respect, again, the Smedley instrument, with its adjustable grip, is a distinct improvement over other forms.

The last objection is not to be seriously considered, first, because hand-grip is one of the most common forms of coördinated movement and is well organized early in childhood, and second, because experience shows that most *S*'s can make their maximal record in three attempts at least.[1]

In any careful or extended investigation *E* must test the calibration of the dynamometer occasionally. For this purpose the instrument is held securely in a vise or other support, and a series of weights are hung upon the stirrup while the scale-readings are compared with the actual weighing (Wallin, p. 64).

For many purposes it is desirable to combine strength of grip with strength of back and strength of legs by adding the data secured in these three tests.

The advantages and disadvantages of using a series of grips in place of a single one are discussed in Test 9.

REFERENCES

(1) A. Binet and N. Vaschide, Expériences de force musculaire et de fond chez les jeunes garcons. *AnPs*, 4: 1897 (1898), 15-63. See also pp. 173-199, 236-252, 295-302.

(2) M. W. Barr, Mental defectives. Phila., 1910.

(3) T. Bolton and Eleanora Miller, On the validity of the ergograph as a measurer of work capacity. *NebraskaUnSd*, 1904. Pp. 79 + 128.

(4) Ada Carman, Pain and strength measurements of 1507 school children in Saginaw, Michigan. *AmJPs*, 10: 1899, 392-8.

(5) J. Clavière, Le travail intellectuel dans ses rapports avec la force musculaire mesurée au dynamomètre. *AnPs*, 7: 1900 (1901), 206-230.

(6) G. E. Dawson, A Study in youthful degeneracy. *PdSe*, 4: 1896, 221-258.

(7) A. Engelsperger u. O. Ziegler, Beiträge zur Kenntnis der physischen u. psychischen Natur des sechsjährigen in die Schule eintretenden Kindes. *EPd*, 1: 1905, 173-235, especially 228-235.

(8) Lucy Hoesch Ernst, Das Schulkind in seiner körperlichen u. geistigen Entwicklung. I Teil. Leipzig, 1906, pp. 78-85.

[1]If, for any reason, *E* considers these sources of error not eliminated, it may be necessary to select a number of *S*'s and coach them in the use of the dynamometer until they can either avoid the errors, or report to *E* when they occur. J. Clavière asserts that to employ only those *S*'s who are thus trained in the use of the instrument is an indispensable condition for successful dynamometry.

(9) W. W. Hastings, A manual for physical measurements. Springfield, Mass., 1902.
(10) A. Hrdlicka, Anthropological investigations of 1000 white and colored children of both sexes. 47th *An. Rep. N. Y. Juvenile Asylum*, 1898.
(11) A. MacDonald, Experimental study of school children. Reprint of chs. 21 and 25, *RepComEd*, 1899.
(12) M. C. Schuyten, Les variations de la force musculaire et le développement intellectuel des élèves; summarized by A. Binet. *AnPs*, 9: 1902 (1903), 448-9, from *Pædologisch Jaarboek*, Ghent, 1902.
(13) M.C. Schuyten, Comment doit on mesurer la fatigue des écoliers? *ArPs*, 4: 1904, 113-128.
(14) F. Smedley, Rept. dept. child-study and pedagogic investigation, reprint from 46th *An. Rep. Brd. Educ. Chicago*, 1899-1900. Also *RepComEd*, 1902, i., 1095-1115.
(15) F. Smedley, do., No. 3, 1900-1901. Also *RepComEd*, 1902, i., 1115-1138.
(16) J. E. W. Wallin, Experimental studies of mental defectives. *EdPsMon*, No. 7, Baltimore, 1912.

TEST 7

Strength of back.—This test, together with the following, has been extensively used in securing an index of the general bodily strength of college students, but has not been applied in most examinations of school children. A fairer index of strength may, however, be gained by its use in combination with strength of grip.

INSTRUMENT.—Back and leg dynamometer (Fig. 17).

METHOD.—*S* stands upon the foot-rest of the instrument, which *E* should then adjust by lengthening or shortening the chain, so that *S*'s body is inclined forward at an angle of about 60 degrees (Fig. 11). *S* should then take a full breath and give a hard lift, mostly with the back and without bending the knees. Two or three trials may be recorded, and the best record used subsequently in computation.

RESULTS.—(1) On the use of this and the succeeding test, with quantitative results as obtained in college gymnasiums, etc., consult Hastings (3), Sargent (4, 5), Seaver (6) and other authorities already cited under anthropometry in general.

(2) Binet and Vaschide (1) found the lift (*force renale*) of 37 boys aged from 12 to 14 years to average 77 kg., with a maximum of 121 kg., and minimum of 56 kg. With 40 young men averaging 18 years of age, the same investigators (2) obtained for the average 146.64, for the maximum 187, and for the mini-

mum 101.6 kg. Hastings (p. 71) publishes measurements of 5000 young men (17-30 years) whose strength of back averages 150.9 kg., *P. E.* 22.1, with a minimal record of 74.5 and a maximal record of 227.3.

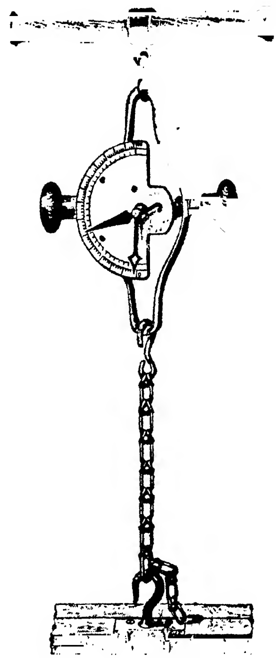


FIG. 17. BACK AND LEG DYNAMOMETER. CAPACITY, 700 KG.

(3) Back lift is roughly about 3.2 times the strength of the right hand.

NOTE.—The adjustment of the chain may, with advantage, be based upon *S*'s height. For this purpose *E* may work out an empirical table of relations between height and the length of chain necessary to give the required position.

REFERENCES

(1) A. Binet and N. Vaschide, Expériences de force musculaire et de fond chez les jeunes garçons. *AnPs*, 4: 1897 (1898), 15-63.

(2) A. Binet and N. Vaschide, La mesure de la force musculaire chez les jeunes gens. *Ibid.*, 173-199.

(3) W. Hastings, A manual for physical measurements for use in normal schools, public and preparatory schools, etc. Springfield, Mass., 1902. Pp. 112.
(4) D. Sargent, Strength tests and strong men at Harvard. *J. Boston Soc. Med. Sci.*, No. 13, 1896-7.
(5) D. Sargent, Anthropometric apparatus, etc. Cambridge, Mass., 1887. Pp. 16.
(6) J. Seaver, Anthropometry and physical examination. New Haven, 1890.

TEST 8

Strength of legs.—This strength test is to be used in conjunction with strength of grip and strength of back. The best records in each of these three tests may be added to secure an index of general bodily strength.

INSTRUMENT.—Back and leg dynamometer.

METHOD.—*S* stands upon the foot-rest of the instrument, with his trunk and head erect and his chest well thrown out, but

FIG. 18. BACK AND LEG DYNAMOMETER, AS USED FOR STRENGTH OF BACK. From D. Sargent, *Anthropometric Apparatus*.

with the knees well bent (Fig. 19). *E* then adjusts the instrument so that the handle, when grasped by *S*, rests against his thighs. *S* should then take a full breath and give a hard lift, mostly with the legs, using the hands to hold the handle in place. Allow two or three trials as before.

RESULTS.—(1) Strength of legs is commonly about 26 per cent. greater than strength of back. Thus, the 5000 men whose records are embodied in Hastings' table have a mean strength of legs of 189.5 kg., *P. E.* 35.3, with a minimal record of 102.2 kg. and a maximal record of 276.8 kg.

(2) The "strength of pull" recorded for Oxford students appears to be somewhat different as regards position of the body from either of the regular American tests with the back and leg dynamometer. Schuster's correlations and those noted in his article (Ref. 11, Test 5) from Pearson's data for Cambridge students probably apply approximately, however, for

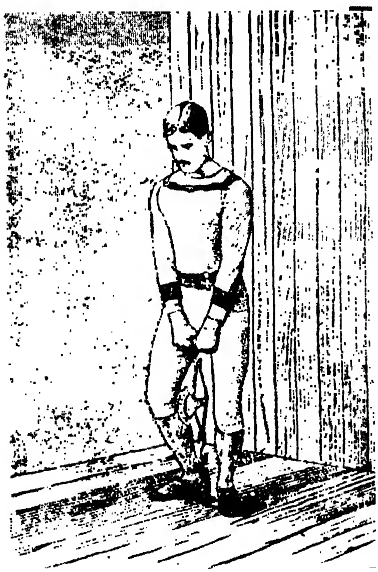


FIG. 19. BACK AND LEG DYNAMOMETER, AS USED FOR STRENGTH OF LEGS. From D. Sargent, *Anthropometric Apparatus*.

Test 7 or Test 8. Correlation of strength of pull with stature was found to be + .21 at Oxford, + .30 at Cambridge; correlation with weight to be + .46 at Oxford and + .56 at Cambridge. At Oxford, also, correlation with vital capacity was found to be + .37.

TEST 9

Endurance of grip.—The object is to test the capacity of *S* to exert maximal muscular exertion, not in a single effort, as in Tests 6, 7 and 8, but during a period of one minute: the test is thus virtually identical with the endurance tests commonly undertaken by means of the ergograph.

Since Mosso's studies of muscular fatigue (32), the ergograph has been extensively employed, not only by physiologists, but also by psychologists and by investigators of school children. The form of the apparatus and the conditions of the test have been widely varied and the numerous factors which affect the test have been exhaustively discussed. In general, the purposes for which the ergograph test has been employed may be summarized thus: (1) to study the physiology of muscular contraction; (2) to detect the presence and to examine the nature and extent of muscular fatigue; (3) like the strength tests, to gain an index of physical capacity or endurance under varying conditions, *e. g.*, as affected by stimulants, narcotics, poisons, exercise, varying diets, etc.; (4) to secure an index of right-handedness; (5) to discover whether physical fatigue is general or local; (6) to discover how mental work affects physical capacity, and, in particular, whether mental fatigue is reflected in muscular fatigue with such clearness that its existence and degree may be ascertained by the examination of some restricted group of the muscles, and (7), on the assumption that physical capacity does measure directly the condition of mental efficiency, to determine the so-called diurnal 'course of power.'

As just intimated, the question of the applicability of the ergograph to these varied purposes raises a large number of problems, in particular that of the nature of fatigue. Since

mental fatigue is reducible to physiological fatigue, the main point at issue is that of the precise nature of the latter.[1] In general, it may be said that, while the relations between mental fatigue and muscular energy are still obscure, we may hope, in principle, to secure some index of the former by our measurements of the latter. If the ergograph is to be employed for exact laboratory experimentation, there is no doubt but that the instrument must be of elaborate construction and the technique equally refined. The development of the ergograph itself from the relatively simple instrument of Mosso to such a complicated apparatus as that devised by Bergström is symptomatic of this gradual refinement of method and progressive analysis of the modifying factors.[2]

But while such problems as the merits of spring vs. weight loading, the relative capacity of a muscle working by isometric and by isotonic contractions, or the most reliable method of isolating the working muscle may be of paramount importance for laboratory investigation, it does not seem, on this account, absolutely impossible, as some writers assert, to secure valu-

[1]On the nature of general vs. local fatigue, on the relative fatiguability of muscle, nerve fiber, nerve cell and synapse, on the distinction between fatigue and exhaustion, and on the nature of other factors, like practise, habituation, readiness for work, warming-up, spurt, etc., that complicate the determination of the fatigue curve, consult Aars and Larguier, Bergström, Bettman, Ellis and Shipe, Hough, Kemsies, Kraepelin, Lee, Lindley, Loeb, Müller, Offner, Rivers, Sherrington, Thorndike, Weygandt and Yoakum. Lee, Offner and Yoakum are perhaps most useful for a preliminary survey. The discussion of these problems which the author attempted in the first edition of this *Manual*, which is here omitted for want of space, may also be consulted. It is perhaps worth repeating here, however, that to understand the nature of mental fatigue and to distinguish between general and local fatigue, it is helpful to separate the objective fatigue (*Ermüdung*), *i. e.*, actual functional inefficiency, from 'weariness' (*Müdigkeit*) *i. e.*, the subjective experience of ennui, loss of interest, or disinclination to work. Thus, it is weariness rather than fatigue which disappears when one's occupation is changed: weariness is fluctuating, uncertain and largely dependent upon the general conditions under which work is being done, while fatigue increases more or less steadily and progressively during our waking moments. If weariness is often specific, fatigue may be more often general and operative to reduce the available energy for work (*Leistungsfähigkeit*) in any direction.

[2]Consult also Cattell (11), Franz (15), Binet and Vaschide (5), Binet and Henri (4), Bolton and Miller (10), Hirschlaff (17), R. Müller (34),

able results for comparative purposes from large numbers of subjects by the use of simpler apparatus and less rigorous technique—provided, of course, that the conditions of experimentation are kept as constant as possible for different subjects and for the same subjects at different times.[1] For this reason the test which is here described is suggested as a practical substitute for the more cumbersome and complicated ergograph.

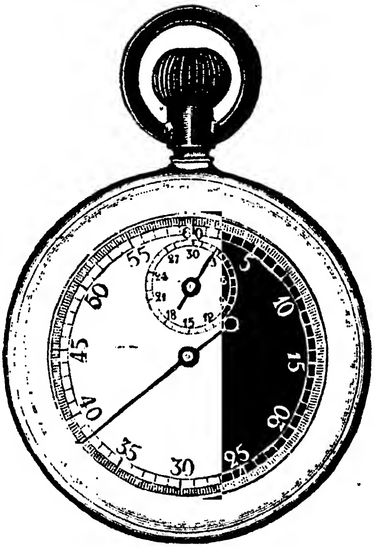


FIG. 20. STOP-WATCH.

APPARATUS.—Smedley dynamometer (Fig. 14). Stop-watch (Fig. 20). Metronome (Fig. 21). [If desired, a kymograph (Fig. 22), with drum support (Fig. 23), a Marey tambour (Fig. 24) and other accessories, or the Mosso ergograph (Fig. 25).]

Two methods are described: one calls for a single, continuous contraction, the other for a series of separate contractions.

[1] In general, it may be expected that minor variable errors will, in a sufficiently long series of tests, be distributed according to the law of chance. Possibly, some portion of the dispute concerning the value of ergograms arises from the fact that certain experimenters have worked upon large numbers of subjects, while others have contented themselves with curves obtained from a single individual.

A. WITH CONTINUOUS CONTRACTION

METHOD.—Set the metronome at 60, *i. e.*, so that it beats once per second. Adjust the dynamometer to *S*'s hand, as in Test 6. Move the friction (recording) pointer of the instrument well over to the right, off the face of the scale. Instruct *S* that he is, at the signal 'now,' to grip as forcibly as possible, to maintain this grip with his utmost effort until told to stop at the end of one minute, and to keep his eyes fixed upon the pointer, so as to hold it as high as possible. (This instruction is designed to act as an incentive to maximal exertion.) Let the instrument be held in the vertical plane with the right-hand edge resting on the table before which *S* is seated.

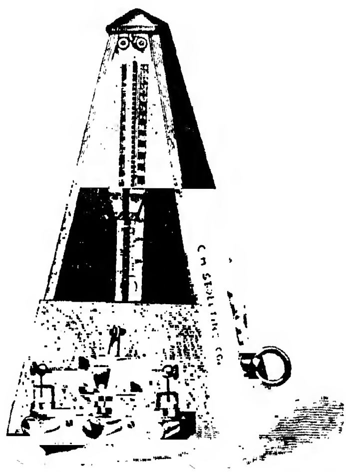


FIG. 21. METRONOME, WITH MERCURY CUPS FOR ELECTRIC CONTACT.

E starts the metronome, and, when he has caught the rhythm, starts the stop-watch, at the same instant saying 'now' for *S* to begin. *E* immediately takes the first reading, and thereafter glances at the scale at every fourth beat of the metronome. In the intervals he records, of course, the reading of the pointer just obtained, estimating to the nearest half-kilogram. If the first reading is secured promptly, *E* will have 16 readings at the end of one minute.

E must practise his work until it becomes automatic. He must take care to keep his eyes directly over the pointer to prevent the error of parallax in reading. For this reason it will be found most convenient for *S* and *E* to sit on opposite sides of the table.

If *S* tends to relax his effort between readings and to exert correspondingly greater effort just as *E* takes the reading, he must be warned against this habit, and urged to exert maximal effort continuously. Even when he does this, the pointer is apt to fall by a series of sudden drops, or even at times to rise as *S* makes a momentary recovery. *E* must take the reading precisely on the beat of the metronome, regardless of the position of the pointer just before or just after the beat.

The stop-watch is used both to test the metronome and to check up the duration of the experiment, but ought to be virtually unnecessary after *E* has practised the experiment sufficiently.

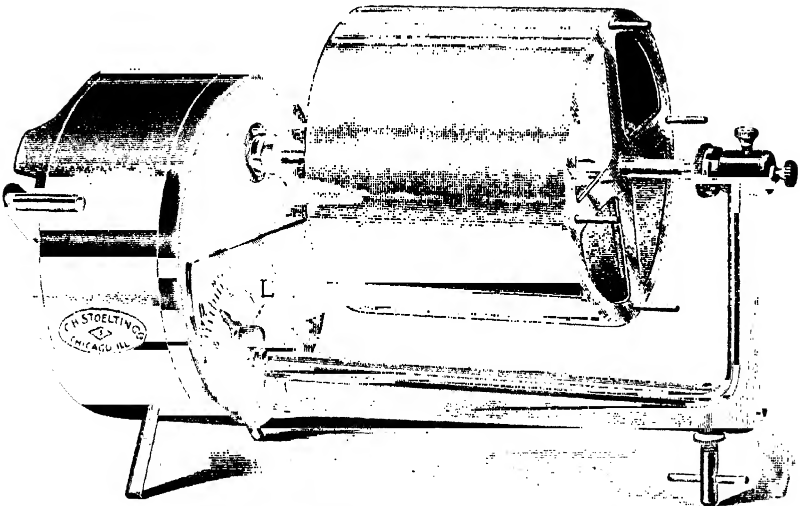


FIG. 22. KYMOGRAPH, IN HORIZONTAL POSITION.

A clock-work mechanism, with regulating fans, in the base, rotates the drum at constant speed and at any desired rate from one revolution in ten seconds to one revolution in ten minutes. Used for making graphic records upon smoked paper.

VARIATIONS OF METHOD.—With the aid of the kymograph described in Test 10, *E* may secure a graphic record of *S*'s work, either by the use of the pneumatic tambour (the Smedley instrument is fitted for pneumatic transmission), or by the use of a simple system of levers to magnify the movement of the handle.[1] The quantitative evaluation of the resulting curve

[1]For a cut showing the method of securing a dynamograph record, see MacDonald (29, p. 1184).

may then be obtained by a series of measurements of the ordinates taken at regular distances and checked by the record obtained as prescribed, or by measurement with a planimeter of the area enclosed by the curve and its base line.

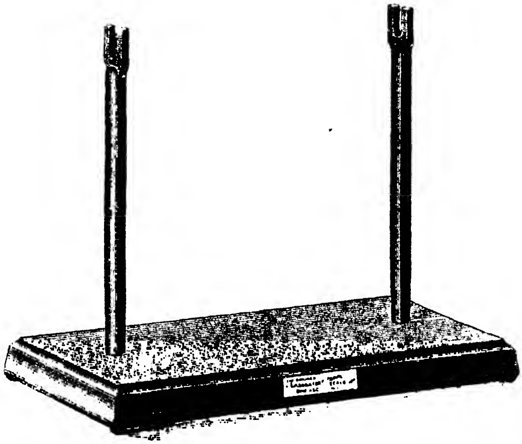


FIG. 23. SUPPORT FOR KYMOGRAPH DRUM WHILE BLACKENING PAPER.

TREATMENT OF RESULTS.—(1) For some purposes the results may be treated by simply averaging the 16 readings, but (2) it will usually be more instructive also to compare the initial with the final stages, in order to secure an index of endurance, or conversely, of fatigue. For this purpose average the first four readings and the last four readings; subtract the latter from the former, and divide the remainder by the average of the first four readings. This may be expressed by the fomula $x = \frac{r_1 - r_2}{r_1}$, when $x =$ the desired index of fatigue, $r_1 =$ the *M* of the first, and r_2 the *M* of the last four readings. Or (3) more simply, one may indicate endurance by the relation of the average to the maximal grip. (4) Following Binet and Vaschide (7), the records of strong, average and weak *S*'s (judged by their maximal grip) may be collated and treated in three groups, in order to trace the presence of the three types of endurance (see below).

B. WITH SEPARATE CONTRACTIONS

METHOD.—Adjust the metronome, dynamometer handle and pointer as in the first method. Inform *S* that, as the word is given, he is to make a series of 16 grips, each as forcibly as possible, and that these grips will be signalled at 4-sec. intervals. *E* then signals 'now' on every fourth beat of the metronome, and takes the readings as previously described.

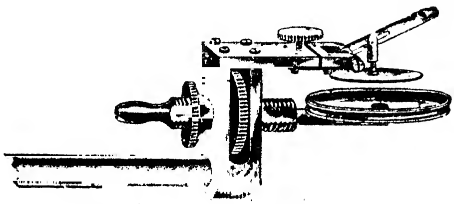


FIG. 24. MAREY TAMBOUR.

For securing tracings by pneumatic transmission. The rubber membrane is not shown.

To hasten the acquisition of skill in conducting this form of the test, *E* will find it helpful to accent the spoken 'now' and to get the swing of the four-beat rhythm by mentally counting the other beats, thus: "*Now,* two, three, four: *now,* two, three, four." As soon as the utterance becomes automatic, *E* can give his whole attention to the readings and the recording of them, and an accurate record can be obtained from very quick and brief excursions of the pointer. Incidentally, some *S*'s may be found who are inclined to hold the pointer up too long: they must be cautioned against this, otherwise fatigue will set in very rapidly.

VARIATIONS OF METHOD.—Substitute the kymograph tracing as suggested above. If this is done, there is no reason why the rate of effort may not be increased, so as to secure 60 or 120 contractions per minute, with a correspondingly more rapid onset of fatigue.

If it is desired to compare results obtained by the dynamometer with those obtained by the common form of ergographic experiment, it is suggested that *E* repeat the experiments made upon Chicago school children. For this purpose substitute the Mosso ergograph for the dynamometer; use the

kymograph for securing the graphic record, and the record furnished by the endless tape, multiplied by the weight, for the quantitative result. Adjust the weight at 7 per cent. of *S*'s weight, and time the contractions to accord with the beats of a metronome set at 30, so as to secure 45 lifts in 90 sec.

TREATMENT OF RESULTS.—This may follow the lines already prescribed.

RESULTS.—(1) The measurement of endurance by the use of the dynamometer has been tried by Binet and Vaschide, though

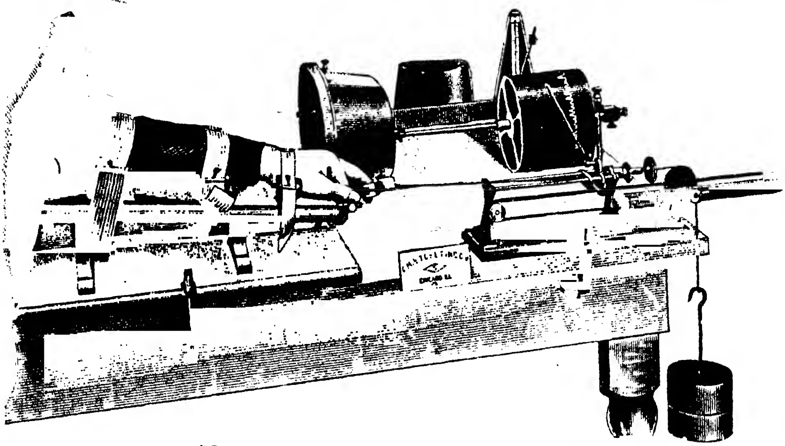


FIG. 25. MOSSO ERGOGRAPH, MODIFIED BY LOMBARD.

under conditions somewhat dissimilar to those we have suggested, upon a group of boys aged 10-13 years (6) and upon a group of young men aged about 18 (7). When five (or ten) grips with each hand, alternately, were required, these authors made out four *types of endurance curve*, viz.: (*a*) a sudden drop, then fairly constant; (*b*) an approximately stationary or constant type, which is quite common; (*c*) a continuous, but gradual drop, and (*d*) a more or less definite rise. The last is rather infrequent (it was not found, *e. g.*, by Clavière, in tests with 15 successive grips), but is sometimes given by vigorous individuals, though the third type is more common for such subjects. Practically 90 per cent. of endurance records can, in the judgment of these writers, be classed in one of these four categories. Table 23 gives average records of five grips, made

with the right hand, by groups representing these four types of endurance curve.

(2) If we accept this hypothesis of 'types,' it is clear that the dynamometer yields a more reliable indication of the comparative muscular capacity when it is employed to test endurance in this way than when merely a single grip is taken, as in Test 6. To take an extreme, though actual, case cited by Binet and Vaschide, it will be seen (Table 24) that, if two subjects

TABLE 23

Types of Endurance in Dynamometer Trials: kilograms (Binet and Vaschide)

	TYPE *a*	TYPE *b*	TYPE *c*	TYPE *d*
1st Grip	23.00	18.70	24.12	17.33
2d Grip	18.45	18.60	22.50	17.70
3d Grip	19.00	19.20	21.17	18.67
4th Grip	18.60	19.40	21.33	18.67
5th Grip	18.20	17.80	19.80	20.67

belong to opposing types, their actual capacities may be completely unsuspected when but a single test is taken.

If we turn to the use of the ergograph, we find the following important, though too often conflicting results:

TABLE 24

Opposed Types of Endurance, 10 Readings (Binet and Vaschide)

NUMBER OF GRIP	1	2	3	4	5	6	7	8	9	10
Subject B	36	34	34	30	29	28	25	26	26	29
Subject R	36	27	42	43	45	42	42	45	46	45

(3) Ergograph curves are affected by *practise* improvement, which, according to Bolton and Miller (10), results (*a*) from 'inurement,' *i. e.*, a fairly rapid "process of hardening and toughening of the skin where it comes in contact with the apparatus and of habituating the muscles to the strains which the unusual effort imposes; (*b*) from improved coördination in the

movements concerned, particularly seen in the disappearance of useless movements; (*c*) from improvement in the rhythmic execution of the contraction, and (*d*) from a slow increase in endurance proper, primarily in the nerve centers. This increase of practise, as Oseretzkowsky and Kraepelin (36) have shown, affects both the height and the number of lifts, and gradually becomes less and less noticeable as maximal practise is attained.

(4) The amount of work that can be done by the muscle is increased if the *rate of lifting* is increased from 30 to 60 or 120 lifts per minute (Oseretzkowsky and Kraepelin).

(5) The work done is conditioned by the *load lifted* or tension of the spring. One cannot, without caution, compare ergograms made with different loads.

(6) The *total amount of physical work* done, as measured by weight × distance, cannot be regarded as a necessarily correct index of the physiological capacity of the muscle; thus, 100 lifts of 25 mm. each may not be assumed to be physiologically equal to 50 lifts of 50 mm. each (Binet and Henri, 4; Franz, 15).

(7) The *weight ergograph* is not adapted to the measurement of muscular capacity (Binet and Henri, 4), hence "the fatigue curves obtained by Mosso and later investigators with weights do not represent the true state of the neuro-muscular system" (Franz, 15).

(8) "The *isotonic* use of a weight or a spring for measuring muscular force is not justified, because two variable factors, extent and force, are introduced," so that an isometric spring (such as the dynamometer) should be used for all comparative experiments (Franz).

(9) With improperly contrived apparatus or inexperienced subjects the ergographic tracing is very liable to be affected by the play of *muscles* other than those under examination (Binet and Henri, 4; Bergström, 2). Müller (34) considers the failure properly to isolate the muscle a fundamental defect of the ergograph.

(10) In addition to these specific criticisms, more general *conclusions of a negative character* may be quoted. Thus, Bol-

ton (8) asserts that the ergograph is not adapted for measuring the degree of fatigue in school children; Bolton and Miller (10) conclude that ergograph records "have slight validity until inurement has become thorough and coördination complete; that the ergograph is quite unadapted to the obtaining of exact statistics upon a large number of individuals, and that records taken upon unpractised subjects, both before and after operations whose influences are thought to affect muscular power, are without the slightest claim to trustworthiness." Similar conclusions are reached by Ellis and Shipe (14) after a retrial of the methods of Keller and of Smedley, also by Thorndike (42), and, with some qualifications, by Bergström (2).

(11) The *effect of physical work* upon ergographic curves seems to vary with the physical condition of the individual and with the nature and duration of the exercise. Thus, Bolton (8, 9) found his ergograms decreased by a two-hour walk, but Oseretzkowsky and Kraepelin found that a one-hour walk caused at first a transient improvement, then a reduction, the first of which they attribute to the increased excitement of central motor tracts, and the second to the damping influence of general muscular fatigue. Smedley (39) tested Chicago children before and after a 40-minute class exercise in the gymnasium with the result that the stronger pupils were little affected, whereas weak and nervous pupils were decidedly exhausted. From this study he concluded that the classes in physical culture should be graded on a physical, instead of on an intellectual basis.

(12) Extensive study of the *effect of mental work* on physical endurance has so far yielded but discordant results. Some of this work, *e. g.*, that of Keller (20), may be thrown out of court at once as careless in plan and execution and merely illustrative of blind infatuation for the ergograph. Typical conclusions of other investigators are as follows: Larguier (24) reports that two hours of mathematics, and Bolton that two hours of adding, definitely increase the ergograph record; Clavière (13), on the other hand, reports that two hours of intense mental work produces a definite and proportionate diminution

of muscular force, whereas intellectual work of medium intensity does not produce any appreciable weakening of endurance. He further confesses his inability to determine the relative fatigue-effect of various school studies. Miss Martyn (31) found, in general, that no consistent decrease in muscular capacity accompanied the fatigue of a school day. The careful ergographic tests of Oseretzkowsky and Kraepelin show that work capacity is increased after one hour of simple addition or learning of 12-place numerals, but that it is lessened if the mental work is rendered more difficult, as by adding under distraction. In an extensively quoted study, Kemsies (21) reports the results of a long series of tests upon a selected group of average, industrious boys who had been trained to the use of the instrument, from which he concludes (*a*) that the ergograph is a reliable indicator of true fatigue (lowered fund of energy as distinct from weariness); (*b*) that subjective feelings of bodily or mental condition may not accord with real capacity; (*c*) that some of the pupils in the Berlin schools show, at least for the time being, signs of overwork; (*d*) that special attention should be paid to pupils who fatigue easily; (*e*) that one can determine for each study its special fatigue-value or 'ergographic-index,' more particularly, that the several studies range themselves, in order from highest to lowest fatigue-index, as follows: gymnastics, mathematics, foreign languages, religion, German, science and geography, history, singing and drawing.

In the attempt to explain these divergences Binet and Henri (4) suggest that we must always distinguish between mental work conducted without emotion and that conducted with emotion. They conclude that the former, if prolonged, may be expected to lessen endurance; the latter to produce a transient increase, followed by a decrease. Kraepelin (23), somewhat similarly, concludes that, while hard mental work certainly reduces muscular energy, deviating results may appear in ergograms on account of the condition of excitement (*Anregung*) that normally accompanies mental work, and that may be expected to affect, either positively or negatively, the tracing which follows such work. Kraepelin further calls attention,

as do Ellis and Shipe, Bergström, Franz and others, to the very large normal variation in the curves of any individual, due to the operation of numerous constant and variable factors, often little understood. Many results are valueless (*e. g.*, in his opinion, those of Kemsies) because of the failure properly to eliminate or evaluate these factors.

(13) The investigations of Christopher (12) and Smedley (39) at Chicago indicate a thoroughgoing correlation between *endurance and class standing*, according to the method of percentile grading, the method of distribution of 12-year-old pupils, and the method of comparison of the endurance of children at and above grade with that of children below grade at each age. Again, boys in the school for incorrigible and truant children were found to exhibit, at every age, less endurance (62 per cent. to 82 per cent.) than normal boys of the same age.

(14) The *endurance of boys* is greater than that of girls at all ages, and the difference becomes very striking during adolescence (39).

(15) The *development of endurance and that of vital capacity* bear a decided resemblance to one another, whether pupils are examined singly or collectively (39).

(16) The *diurnal 'course of power'* according to the Chicago experiments may be expressed as follows: "(*a*) The extremes of endurance and fatigue in school are greater in the morning than in the afternoon; (*b*) a higher grade of power is found in the morning session in children attending two sessions daily; (*c*) while endurance is not as great, it is better sustained in the afternoon." Compilations of the ergograms of 1127 pupils place the maximum at 9 A. M. and the minimum at 12 noon. Kemsies considered the first two morning hours the best. Experiments upon adults by Lombard (28), Harley (16), Storey (41) and Marsh (30) exhibit considerable lack of agreement with one another or with the Chicago results, though Marsh summarizes them by the statement that the curve of strength efficiency seems well established for the following course: "a beginning minimum in early morning, a fairly rapid rise till 11, a level or slight decline till 1 P. M. ($\pm$ 1 hour), an increase to the maximum at 5 P. M. ($\pm$ 1 hour), thence a fall till bedtime."

(17) Kemsies concludes that Monday and Tuesday, or the first two days after any rest pause, are the *best days for general efficiency,* and he further concludes that vacations exert a powerful effect upon efficiency; but, since this effect cannot be traced for longer than four weeks, school terms should be broken up by more frequent vacations of shorter duration.

(18) If ergographic contractions are continued to the point of *exhaustion,* we have both the sum-total of the height of the lifts and their number for indexes of the neuro-muscular condition. Hoch and Kraepelin (18) are of the opinion that, in this case, the height of contraction is conditioned by the state of the muscles, but the number of contractions by the state of the central nervous system; the two factors should, therefore, be reported separately for their diagnostic value. On the other hand, Lombard (28) concludes that, at least when the contraction is not faster than once per second, the amount of fatigue experienced by the central nervous system does not correspond to the number of lifts, but rather to the strength of the motor impulses discharged, so that the sum-total of the height of lifts is the more accurate index of the state of the central nervous mechanism.

(19) According to Lombard, endurance is increased by exercise, rest (especially sleep), food, increased atmospheric pressure and by small doses of alcohol, but lessened by general and local fatigue, hunger, lessened atmospheric pressure, high temperature, especially with high humidity, and by tobacco. Oseretzkowsky and Kraepelin find that coffee increases the height of lifts, and that alcohol, in quantities from 15 to 20 g., causes at first a considerable increase, especially in the number of lifts, but that this soon disappears. On the other hand, Rivers and Webber (38) have discovered that small doses of alcohol (5-20 cc.) fail to produce any appreciable modification of the ergographic record if proper precautions are taken to keep the subject in ignorance as to when alcohol is administered. The results of previous workers are, therefore, presumably due to the influence of other factors, particularly interest and sensory stimulation, and no future work on the effects of small doses of alcohol can be acceptable unless these factors are controlled.

Harley (16) concludes that "moderate smoking, although it may have a slight influence in diminishing the power of doing voluntary muscular work, neither stops the morning rise nor, when done early in the evening, hinders the evening fall."

REFERENCES

(1) K. Aars and J. Larguier, L'effort musculaire et la fatigue des centres nerveux. *AnPs*, 7: 1900 (1901), 187-205.

(2) J. A. Bergström, A new type of ergograph, with a discussion of ergographic experimentation. *AmJPs*, 14: 1903, 510-540.

(3) S. Bettmann, Ueber die Beeinflussung einfacher psychischer Vorgänge durch körperliche u. geistige Arbeit. *PsArb*, 1: 1895, 152-208.

(4) A. Binet and V. Henri, La fatigue intellectuelle. Paris, 1898. Pp. 336.

(5) A. Binet and N. Vaschide, Examen critique de l'ergographe de Mosso. *AnPs*, 1897 (1898), 253-266. Also, Un nouvel ergographe, dit ergographe à ressort. *Ibid.*, 303-315.

(6) A. Binet and N. Vaschide, Expériences de fòrce musculaire et de fond chez les jeunes garçons. *Ibid.*, 15-63.

(7) A. Binet and N. Vaschide, La mesure de la force musculaire chez les jeunes gens. *Ibid.*, 173-199. See also *ibid.*, 236-244, 245-252 and 295-302.

(8) T. Bolton, The reliability of certain methods for measuring the degree of fatigue in school children. *PsR*, 7: 1900, 136-7.

(9) T. Bolton, Ueber die Beziehungen zwischen Ermüdung, Raumsinn der Haut und Muskelleistung. *PsArb*, 4: 1902, 175-234.

(10) T. Bolton and Eleanora Miller, On the validity of the ergograph as a measurer of work capacity. *NebraskaUnSd*, 1904, 79 + 128.

(11) J. Cattell, An ergometer. *Sci*, n. s. 5: 1897, 909-910.

(12) W. Christopher, Rept. on child-study, reprint from *An. Rep. Brd. Educ. Chicago*, 1898-99.

(13) J. Clavière, Le travail intellectuel dans ses rapports avec la force musculaire mesurée au dynamomètre. *AnPs*, 7: 1900 (1901), 206-230.

(14) A. C. Ellis and Maud Shipe, A study of the accuracy of the present methods of testing fatigue. *AmJPs*, 14: 1903, 496-509.

(15) S. I. Franz, On the methods of estimating the force of voluntary muscular contraction and on fatigue. *AmJPhg*, 4: 1900, 348-372.

(16) V. Harley, The value of sugar and the effect of smoking on muscular work. *JPhg*, 16, 1894, 97-122.

(17) L. Hirschlaff, Zur Methode und Kritik der Ergographenmessungen. *ZPdPs*, 3: 1901, 184-198.

(18) A. Hoch and E. Kraepelin, Ueber die Wirkung der Theebestandtheile auf körperliche und geistige Arbeit. *PsArb*, 1: 1896, 378-488.

(19) T. Hough, Ergographic studies in neuro-muscular fatigue. *AmJPhg*, 5: 1901, 240-266.

(20) R. Keller, Pädagogisch-psychometrische Studien. *BiZb*, 14: 1894, 24-32, 38-53, 328-336.

(21) F. Kemsies, Zur Frage der Ueberbürdung unserer Schuljugend. *DMdW*, July 2, 1896, 433. See also his Arbeitshygiene der Schule auf Grund von Ermüdungsmessungen. Berlin, 1898, and *SmAbPdPs*, 2: 1899, Heft 1., pp. 64.

(22) E. Kraepelin, Zur Ueberbürdungsfrage. Jena, 1897. Pp. 49.

(23) E. Kraepelin, Ueber Ermüdungsmessungen. *ArGsPs*, 1: 1903, 9-30.

(24) J. Larguier, Essai de comparison sur les différentes méthodes proposées pour la mesure de la fatigue intellectuelle. *AnPs*, 5: 1898 (1899), 190-201.

(25) F. S. Lee, Fatigue, in the Harvey Lectures, Phila., 1906, 169-194. Also *J. Amer. Med. Ass.*, 46: 1906, 1491, and in *Studies in Physiology*, Columbia Univ., 1902-7.

(26) E. H. Lindley, Ueber Arbeit und Ruhe. *PsArb*, 3: 1900, 482-534.

(27) J. Loeb, Muskelthätigkeit als Mass psychischer Thätigkeit. *ArGsPhg*, 39: 1886, 592-7.

(28) W. P. Lombard (a), The effect of fatigue on voluntary muscular contractions. *AmJPs*, 3: 1890, 24-42. (b) Some of the influences which affect the power of voluntary muscular contractions. *JPhg*, 13: 1892, 1-58.

(29) A. MacDonald, Experimental study of school children, etc., reprint of chs. 21 and 25, *RepComEd*, 1899.

(30) H. D. Marsh, The diurnal course of efficiency, Columbia Univ. diss., N. Y., 1906. Pp. 99.

(31) Gladys Martyn, The evidence of mental fatigue during school hours. *JEPd*, 1: 1911, 137-147.

(32) A. Mosso, Fatigue. Eng. tr., N. Y., 1904. Pp. 334.

(33) G. E. Müller, Review of A. Waller's "The Sense of Effort." *ZPs*, 4: 1893, 122-138.

(34) R. Müller, Ueber Mosso's Ergographen, etc. *PhSd*, 17: 1901, 1-29.

(35) M. Offner, Mental fatigue. Eng. trans. by G. M. Whipple, Baltimore, 1911. (Useful bibliography).

(36) A. Oseretzkowsky and E. Kraepelin, Ueber die Beeinflussung der Muskelleistung durch verschiedene Arbeitsbedingungen. *PsArb*, 3: 1901, 587-690.

(37) W. Rivers and E. Kraepelin, Ueber Ermüdung u. Erholung. *PsArb*, 1: 1896, 627-678.

(38) W. Rivers, The influence of small doses of alcohol on muscular activity. *PsBu*, 5: 1908, 49. See also Rivers and H. Webber, The influence of small doses of alcohol on the capacity for muscular work. *BrJPs*, 2: 1908, 261-280.

(39) F. Smedley, Rep. dept. child-study and pedagogic investigation, reprint from 46th *An. Rep. Brd. Educ., Chicago*, 1899-1900. Also *Rep ComEd*, 1902, i., 1095-1115.

(40) T. Storey, The influence of fatigue upon the speed of voluntary contraction of human muscle. *AmJPhg*, 8: 1903, 355.

(41) T. Storey (a), Some daily variations in height, weight and strength. *Am. Phys. Ed. Rev.*, 6: 1901. (b) Daily variation in the power of voluntary muscular contraction. *Ibid.*, 7: 1902. (c) Studies in voluntary muscular contraction, Stanford Univ. Press, 1904.

(42) E. L. Thorndike, Mental Fatigue. *PsR*, 7: 1900, 466-482, 547-579.

(43) W. Weygandt, Ueber den Einfluss des Arbeitswechsels auf fortlaufende geistige Arbeit. *PsArb*, 2: 1897, 118-202.

(44) C. S. Yoakum, An experimental study of fatigue. *PsMon*, 11: 1909, No. 46. Pp. 131. (Bibliography of 108 titles).

TEST 10

Quickness or rate of movement: Tapping.—This has probably been more frequently applied than any other 'motor test,' and has been thought to afford a better index of motor capacity than any other single test. Recent work with tapping, however, while not discouraging the belief that the test has value, has shown that we cannot regard speed of voluntary movement as an unequivocal and comprehensive 'index of voluntary motor ability,' because a high gross rate does not necessarily go hand in hand with high speed in other phases of motor response, and because, moreover, we do not know precisely what may be the ultimate neural or psychophysical factors that condition the rate.[1]

Aside from its use in the attempt to secure this 'index of voluntary motor ability,' the tapping test has been employed to secure an index of right-handedness (for which purpose it may be advantageously combined with Tests 6, 9, 11 and 12), and to secure an index of fatigue (likewise preferably in conjunction with other tests of physical capacity). These several indexes have been studied in various comparative investigations, more especially in estimating sex and age factors in motor development and the relation of physical to mental ability at large.

The duration of the test has ranged from 10 sec. (Abelson), or 15 sec. (Burt), to one minute, or to the completion of 300 taps (Dresslar), and the method from the simple making of dots or vertical marks with pencil and paper (Binet and Vas-

[1]To quote from Wells (26, p. 444): "What is the precise physiological significance of the maximum rate is by no means well made out. * * * It seems to be generally conceded that it is limited by the refractory phase of the synapses in the motor pathways, but that does not make the tapping test a measure of the period of this refractory phase; at least not in the earlier stages of practice. * * * In the beginning, as we ordinarily have to apply the test, the factors in speed are probably those of coördination mainly, and cannot be expected to afford information about the condition of the motor pathways as given in the refractory phase."

On the other hand, the correlations obtained by Abelson with 'special class' children convince him that this test has for them a high diagnostic value. He says: "Evidently tapping, which is quite a mechanical procedure for normal children, requires from defective children a special effort of attention and will."

chide), or the puncturing of a small square of paper with a pointed stylus (Abelson, Burt) to the execution of difficult trilling movements upon telegraph keys.[1] The apparatus here prescribed is somewhat elaborate, but experience has shown that the tapping test cannot be conducted to the best advantage without careful control of experimental conditions and the use of a reliable device for recording a continuous and permanent record, such as the graphic method supplies.

MATERIALS.—Tapping board, 55x10 cm., with brass plates 10 cm. square on either end (Fig. 26). Tapping stylus, with flexible connecting wire attached. Kymograph (Fig. 22), with accessories—paper, smoking device, shellac solution. Double time-marker (Fig. 27). Seconds' pendulum (Fig. 28) or other

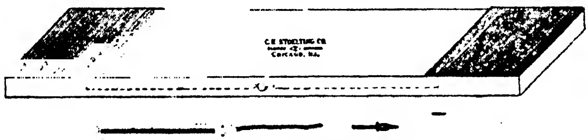


FIG. 26. TAPPING-BOARD.

noiseless instrument arranged to give electric contacts once per sec. Support with levelling screw and right-angle piece to hold time-marker. Table clamps for tapping board. Large sheet of gray or white cardboard. Suitable supports and clamps for holding cardboard. Two short-circuiting keys (Fig. 29), or simple knife switches. Stop-watch. Four dry or Leclanché cells. Flexible covered wire with connector tips or ordinary No. 18 annunciator wire. [A swivel chair, adjustable in height, and an ammeter or battery-tester are also convenient, though not absolutely essential.]

PRELIMINARIES.—(1) Clamp or screw the tapping board securely at the edge of a table in such a manner that *S* may

[1] Tests like rapid counting aloud or the rapid reading of digits or the reaction-time test are not psychologically comparable to the tapping test. Again, the form of test used at Columbia University and elsewhere to measure rate of movement (making a dot as rapidly as possible in each of 100 one-cm. squares) is not equivalent to the tapping required in most quickness tests, since a certain amount of precision is demanded of each movement, and that test therefore stands midway between Tests 10 and 11, as here prescribed. The so-called 'tapping' test used by Yoakum and by Mrs. Squire to investigate fatigue is not a test of motor coördination, but rather of attention and memory.

have free access to either end of the board for using either right or left hand. Arrange *S*'s chair so that he sits sidewise to the table with his forearm resting comfortably along the tapping board and his hand directly over the metallic plate.

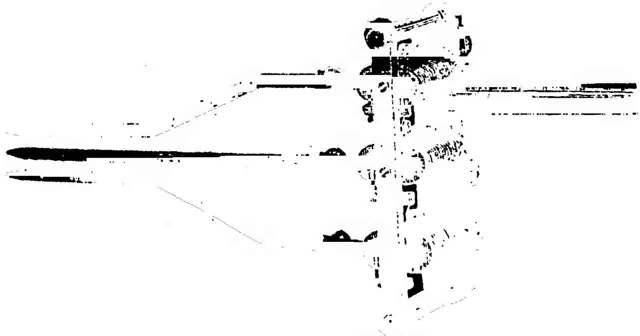


FIG. 27. TRIPLE TIME-MARKER.
The double and the single time-marker are of similar construction.

(2) Place the kymograph in a horizontal position, screened from *S*'s view by the sheet of cardboard.[1] Adjust the fans or gear-wheels so that the drum makes (for a 30-sec. test) one revolution in about 40 sec.

(3) Remove the drum and cover with the prepared paper by simply moistening the gummed end, taking care to draw it evenly and tightly around the drum. Blacken the paper by revolving it slowly in a smoky flame.[2] Replace carefully in the kymograph.

[1]The screen is to avoid the distraction of *S*'s attention by the operation of the apparatus. If separate tables are used for tapping board and kymograph, this may not be necessary, but it is commonly more convenient to assemble all the apparatus on a single table.

[2]An oil stove from which the top is removed is excellent for this purpose, as the flame is very sooty and not so hot as the gas flame often employed. A simple support (Fig. 23) is used to hold the drum, both for the smoking and for the subsequent removal of the paper. For this and other details in the use of the kymograph, consult Titchener, *Experimental Psychology*, Vol. I, Part II, pp. 172-180.

If an extended series of tests is to be made, cover the drum permanently with the regular kymograph paper, and, for the records, superpose two narrower strips, say 75 mm. wide. These strips are wide enough to record the right and left-hand efficiency of one *S*: they can then be removed promptly for fixing, and thus the danger of injury is lessened, the ease of handling increased, and the blackening of the metal drum is less likely to be a source of annoyance.

(4) Adjust the time-marker on the support so that the pointers bear upon the drum with just sufficient pressure to make a satisfactory tracing. The pointer must move in a plane parallel to the plane of a tangent drawn through the point of contact.

The manipulation of the apparatus may be facilitated by fastening upon the table, in front of, and parallel to the surface of the drum, a straight bit of wood somewhat longer than the drum. Let the foot of the tripod which contains the levelling screw stand away from the drum, and the other two feet bear against the wooden strip. A half turn of the levelling screw will then free the pointers from the drum, and the entire support with the time-marker may be slid along to a new position, when another half turn of the screw will quickly adjust the pointers for the next record.

(5) Wire one signal-magnet in series with the tapping board stylus, short-circuiting key and two cells of the battery. The magnet will then be set in motion by the tapping when the key is closed.

(6) Wire the second magnet in series with the pendulum, second short-circuiting key and remaining two cells of the battery. This magnet will then be set in motion by the pendulum when its controlling key is closed, and will thus beat off the time-line.

METHOD.—(1) Seat *S* for the use of his right hand. Instruct him to tap as rapidly as possible from the signal 'now' to the signal 'stop,' which will be given about one-half minute later. Tell him to pay attention only to his tapping. He may be allowed to exercise some latitude with regard to the type of movement used (short or wide excursion), unless he is inclined to adopt a very heavy whole-arm pounding movement. The most favorable movement for most *S*'s is that obtained by resting the elbow on the tapping board and using both the wrist and elbow joints.

(2) Start the seconds' pendulum and close the time-line circuit.

(3) Start the kymograph, and when it is fairly in motion give *S* the signal 'now.'

(4) When *S* has had time to respond, *i. e.*, after two or three sec., throw in the record magnet by closing its key, and at the same instant start the stop-watch.

(5) When a clear record of 30 sec. has been inscribed, stop the watch, break the record circuit, signal 'stop' to *S*, stop the kymograph and open the time-circuit. [*E* must practise the whole series of operations until they run smoothly and automatically.]

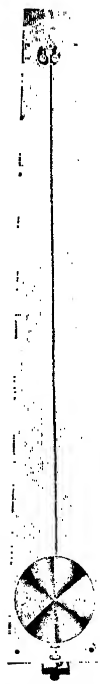


FIG. 28. SECONDS' PENDULUM.

(6) Now adjust the pointers for a new record. Let *S* sit facing in the other direction, and test the left hand by the use of the plate at the other end of the tapping board, following the directions given for the test of the right hand. The interval between the end of the first and the beginning of the second record is preferably kept constant at 30 sec.

VARIATIONS OF METHOD.—(1) The duration of the test may be lengthened to 45 sec. or longer, or shortened to 10 or 20 sec.

It is, however, desirable, for the sake of comparison, that a standard duration be employed. Thirty sec. is adequate for all ordinary purposes.

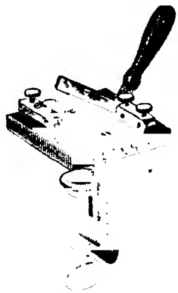


FIG. 29. SHORT-CIRCUITING KEY (DU BOIS REYMOND)

(2) For certain types of investigation, *e. g.*, when disorders of motor innervation are suspected, it is recommended that the plan of experiment suggested by Wells and followed by him and by Strong be adopted. In Wells' terminology each 30-sec. period of continuous tapping is termed one *series,* and is regarded as composed of six 5-sec. *intervals.* Five such series constitute one *record,* and two records, the first usually with the right, the second with the left hand, constitute one standard *experiment.* The hands do not alternate from series to series, but only from record to record. The first one used is called the 'preceding,' the other the 'following' hand. Between each two series there is a rest-pause of 2.5 min., during which *S* should refrain from all muscular effort.

(3) To make the test comparable to the form employed by some investigators, an ordinary 'sending' telegraph key may be substituted for the tapping board and stylus. But the key has the disadvantage of imposing a certain restriction upon the type of movement, and will be found in practise to reduce the record of many *S*'s.

(4) By using the key, *E* may compare *S*'s rate of tapping with different fingers. For this purpose it is well to fasten down the forearm with a strap at the wrist, so as to allow movement with the fingers only.

(5) Again, by using the key, a trilling movement, executed

by alternate movements of the index and middle fingers, may be substituted for the regulation tapping movement. Without practise this movement is quite difficult for some *S*'s, whereas for others, notably for those who have practised trilling exercises on the piano, it is comparatively easy.[1] For this reason this form of experiment is not advised, save for some exceptional purpose, *e. g.*, testing the effect of practise upon the acquisition of a new bit of manual dexterity. Other modifications suggest themselves, such as trilling with the 4th and 5th fingers—an exercise likely to be unfamiliar even to *S*'s who have 'taken lessons.'

TREATMENT OF RESULTS.—(1) When the record has been made, use any pointed article to mark it for future identification (*S*'s name or number, date, hand used, etc.); then remove carefully for preservation. A simple and satisfactory method is to pour a very thin solution of shellac and wood alcohol or of powdered resin (not over 10 per cent.) in alcohol, into a saucer or shallow dish, and to pass the strip through this, smoked side up. Hang the record up to dry, and pour the solution back into a wide-mouthed bottle, where it should be kept tightly stoppered. The record will dry in a few minutes and can then be trimmed and handled with impunity.

(2) The result of the test is commonly expressed simply by the total number of taps executed, but it is quite as important, if not more important, to consider changes of speed during each trial. The requisite data must be secured by the rather laborious process of counting the strokes made by the recording-magnet upon the blackened paper, and tabulating them by 5-sec. intervals, as illustrated below. The 'total efficiency' of a 'record' (5 trials with the same hand) is the average of the sum of the taps per trial. This serves as the gross index of speed: the rates for the 6 intervals within each trial afford an opportunity for studying variations in performance.

[1]The effect of piano practise, as the investigations of Binet and Courtier (3) and of Ralf (20) show, is to improve coördination of movement, *i. e.*, its regularity, smoothness, etc., but not to increase the natural capacity for speed or rate of movement.

The use of an electric counter, such as some investigators have employed, would eliminate this work, but the counter gives no indication of changes in speed during the trial. Moreover, the electric counter is not reliable: even with 10 or 12 cells of battery, it will miss a quick tap which the graphic method will record. It follows that all results based on the use of the counter are to be looked upon with suspicion, so that the conclusions of Bagley, Bolton, Marsh, Kelly, Smedley, and possibly those of Bryan, Davis, Gilbert and Dresslar should be accepted with reservation. If a mechanical counter of the dial type is to be used, preference should be given to the Ewald chronoscope, which, as Dunlap has shown, operates perfectly for the tapping test and is a useful instrument for many other laboratory purposes.

(3) To secure an index of fatigue, *E* may compare the record of the first 5 (or 10) sec. with that of the last 5 (or 10) sec. by the use of the formula for determining the relative loss of efficiency given in Test 9.

Wells has published extensive conclusions concerning fatigue in tapping that are based upon a differently computed index. The average number of taps executed in the 2d, 3d, 4th, 5th and 6th 5-sec. intervals are divided by the number of taps executed in the 1st 5-sec. interval. This index is somewhat misleading, in so far as a high index indicates a low degree of fatigue. If it is deemed worth while to relate the last five to the first of the six intervals to compute fatigue (a procedure which seems less desirable than the one above prescribed), it would be better, in the author's opinion, to subtract the average in question from the initial speed and divide the loss by the efficiency of the first interval.

TABLE 25.

Sample Record of a Tapping Test (Wells)

NUMBER OF INTERVAL	1ST	2D	3D	4TH	5TH	6TH	TOTAL
1st series	41	37	35	34	34	32	213
2d series	41	37	36	35	34	34	217
3d series	40	39	37	37	35	34	222
4th series	40	39	37	36	36	35	223
5th series	41	39	38	37	36	36	227
Averages	40.6	38.2	36.6	35.8	35.0	34.2	220.4

(4) To secure an index of right-handedness, *E* may compute the percentage of the left-hand to the right-hand efficiency. The fatigue-index of the right hand may also be compared with that of the left hand in a similar manner.

Typical Results.—Table 25 shows a sample record of the work of a normal adult with the right hand when near the limit of practise. The tabulation is in accordance with that recommended when five 30-sec. series are recorded.

General Conclusions.—Although the tapping test is one of the most objective that can be applied, and although it has been tried by a large number of investigators (see the references at the end of the test), the results have not been always accordant, and, with the exception of the recent work of Wells, have not been so treated as to afford real insight into the factors that underlie their appearance. The lack of accordance is to be attributed in large part to differences in method of procedure. Differences in apparatus, too, have been sufficient to account for some discrepancy, as has already been pointed out. As regards method, the duration of the test, to instance a single point, has varied from 5 sec. (Binet and Vaschide, Bryan, Kirkpatrick) to 2 min. (Thompson), with intermediate durations, such as 10 sec. (Bagley), 30 sec. (Smedley), 45 sec. (Gilbert), 60 sec. (Kelly), or the test has been conducted in 5 series of 5 sec. each (Bolton).[1]

In so far as these divergences of method may be neglected, we may note the chief conclusions of interest concerning the tapping test as follows:

(1) In general, the *maximal rate* of voluntary movement varies with the individual, with sex, with maturity, with the side of the body used, with practise, with the number of trials (duration of experiment), with fatigue, with mental excitement, with the time of day, but not, within wide limits, with the amplitude of the movement.[2] For practised adults the initial rate is about 8.5 taps per sec. (Dresslar), with a range from 5 to 14 taps.

(2) *Constant individual differences* in rate of tapping can be demonstrated without much difficulty, but we cannot at

[1]The situation here, as in most tests, shows clearly how desirable it would be to establish some standard form of test and to use it alone for all comparative purposes.

[2]For a fuller discussion of these conditions, consult Dresslar, Bryan, Strong, Abelson and Wells.

present explain them, save to say that they are conditioned in a general way by fundamental neural factors, or by these plus differences in ability to coördinate voluntary movement. Thus, in 10 adults tested by Wells (10 series for each hand), the average total efficiency (taps in 30 sec.) was approximately 194, but the fastest *S* averaged 225, and the slowest 153. Since the m. v. is small (here approximately 1 to 3 per cent.), these figures undoubtedly indicate persistent characteristic differences.

(3) *Dependence on age.* The rate of tapping increases with age, at least between 6 and 18 years. The slight drop at 13, upon which Gilbert comments, appears in Bryan's tables with some qualifications, but not so clearly in Smedley's results, which are reproduced herewith: it will be seen, however, that boys make no apparent gain from 13 to 14.

TABLE 26

Dependence of Rate of Tapping upon Age (Smedley)

AGE	BOYS			GIRLS		
	NUMBER TESTED	TAPS IN 30 SECONDS		NUMBER TESTED	TAPS IN 30 SECONDS	
		Rt. Hand	Lt. Hand		Rt. Hand	Lt. Hand
8	31	147	117	31	146	117
9	60	151	127	44	149	118
10	47	161	132	48	157	129
11	49	169	141	48	169	139
12	44	170	145	50	169	140
13	50	184	156	45	178	153
14	40	184	155	67	181	157
15	37	191	169	48	181	159
16	21	196	170	50	188	167
17	13	196	174	40	184	162
18	3	197	183	24	193	169

(4) *Dependence on sex.* The results of most investigators lead to the conclusion that boys are faster than girls, and that this sex difference increases with age. Thus, Miss Thompson, who worked with a small number of adults, found that 88 per cent. of men exceeded the median speed of women. Similarly,

Burt and Moore place the median for boys at 86.5 and for girls at 80.5 taps in 15 sec., and find that 69.8 per cent. of boys exceed the median speed of girls. Bolton, however, has reported that "the girls are uniformly better than the boys," while Bryan found girls superior at 13, when they showed improvement and the boys little or none—a tendency that is apparently allied to the actual crossing of the curves of height and weight. More elaborate experiments upon adults (10 men and 10 women) by Wells (28) now seem to indicate that women surpass men in tapping with the right hand in the first experiment, whereas elsewhere they are inferior: the sex differences found by this investigator are said to be "mainly in those features of the experiment which especially involve the affective factor in the subject's attitude, and they are manifestations of the greater responsiveness of the women to this affective element."

(5) The *index of right-handedness* (per cent. left-hand is of right-hand efficiency) was found by Wells to range from .81 to .94, average .90, for adults, and by Smedley to vary with age in the case of school children in such a manner that the average index was .82 at the age of 9, and .89 at the age of 18. It seems, therefore, as if right-handedness, so far as tapping is concerned, is more pronounced in childhood than in adult life. Wells also states (28) that "the right and left hands are farther apart in women," though the relationship is more variable in them than in men. The index is to a small extent conditioned by the order in which the records are taken. For normal *S*'s each hand tends to be relatively better when it precedes than when it follows.

(6) *Right-handedness and intelligence.* Smedley's conclusion that there exists a positive correlation between degree of right-handedness and school standing, *i. e.*, that the left-hand more nearly approaches the right-hand efficiency in the case of dull and backward pupils, is not confirmed by the results of Bolton.

(7) *Warming-up.* In practically every continuous psychophysical activity there appears a tendency to improvement, due to what the Germans have termed *Anlauf.* This 'warming-up'

is a kind of momentum not identical with practise, and its effect is to increase or heighten the activity, and thus to retard or even to obscure the appearance of indications of fatigue. In tapping, we observe fatigue within each 30-sec. series, but a comparison of successive series within a record will show the improvement due to warming-up. With 2.5 min. rest-pauses, Wells found this factor to be clearly present (up to the 7th series at least) in right-hand records, but by no means so evident in left-hand records. The effect of warming-up appears to be primarily operative in increased immunity to fatigue, and is markedly augmented by practise, *e. g.*, in tests continued for 20 days.

(8) *Spurts.* The curve of performance in tapping, as well as any psychophysical activity, is also liable to be influenced by short periods of increased activity, which, to continue the analogy of the race-track, may be termed 'spurts' (German, *Antriebe*). Thus, Wells' discovery that the first experiment usually excels the second in women, whereas the reverse may be true in men, is referred to a special incitement of novelty (*Neuigkeitsantrieb*), which affects the women markedly. Similarly, each 'record' may be affected by an initial spurt (*Anfangsantrieb*) or by a terminal spurt (*Schlussantrieb*). These dynamogenic factors obviously tend to obscure the real effects of fatigue.

(9) *Fatigue and the fatigue-index.* (*a*) As just stated, the speed of tapping normally declines after the 1st 5-sec. interval, until it is approximately six-sevenths as great in the last as in the 1st interval (Wells). In 45-sec. series the fatigue-index (loss of last 5 sec. divided by initial 5 sec.), according to Gilbert, is highest in young children (24 per cent. at 8 years), and declines thence irregularly to 12.7 per cent. at the age of 15. Tests by the author of fifty 8th-grade grammar-school boys reveal a fatigue-index (ratio of loss in 3d 10 sec. to 1st 10 sec. in 30 sec. tapping) of .137, m. v. .048 for the right, and .15, m. v. .046, for the left hand. Wells' so-called 'fatigue-index' ranges for normal *S*'s between .85 and .95.

(*b*) According to Gilbert, the fatigue-index is higher for boys than for girls, but boys tap faster throughout each trial, so that their net efficiency is higher.[1]

(*c*) Kelly (15), who worked on a small number of children with a "fatigue-counter," found that "A"-grade pupils fatigued less than "C"-grade pupils. His index (the per cent. of the last to the 1st 10 sec. in a 60-sec. trial) was for the former, with the finger 87.2 per cent., with the arm 88.0 per cent.; for the latter, with the finger 77.0 per cent., with the arm 76.4 per cent. In the author's tests no correlation could be discovered between fatigue-index and school standing.

(*d*) The effect of fatigue is progressively to 'level up' individual differences in speed. In other words, individual differences are more evident in initial than in terminal intervals (Bliss, Wells).

(*e*) Objective fatigue (slowing in rate) persists after practise, but the subjective feeling of fatigue may be eliminated thereby.

(*f*) The fatigue induced by 30-sec. tapping is apparently completely eradicated by a 3-min. rest-pause (Wells).

(*g*) The fatigue-index of right and left hands shows only slight correlation (Wells). The author's tests, however, show a correlation in the case of 50 boys of .33. In some persons the left hand is less susceptible to fatigue than is the right hand, though the reverse is the rule.

(*h*) The subjective experience of fatigue, as has been intimated, does not accord with the objective fatigue-loss.[2]

---

[1] It is more reasonable to interpret this higher index of boys as an expression of greater zeal and enthusiasm than to follow Havelock Ellis in his inference that this is an example of the "more continuous character of woman's activity," especially since Miss Thompson, from comparative tests of adults, concludes that men surpass women both in initial rapidity and in power to sustain it, and since other tests show that high initial speed tends to be accompanied by a high fatigue loss.

[2] According to Wells: "The objective fatigue phenomena which we note in the test are in all probability either a fatigue phenomenon in the refractory phase or a lowered efficiency of coördination, especially a product of altered synaptic conditions; the sensations of fatigue, on the other hand, may with equal assurance be ascribed to tissue changes within the muscles that take place as a result of their continued effort. In this test, therefore, the fatigue sensations are absolutely no indications of the actual fatigue conditions, and any traceable correspondence between fatigue sensation and fatigue of performance must be regarded as almost wholly a product of reflex inhibition" (26, p. 473).

(10) *Practise.* (*a*) The effect of practise is to produce a gradual improvement in speed, with, of course, occasional losses.

(*b*) The rise of the curve of efficiency is not, as in most activities, more rapid at the beginning than elsewhere.

(*c*) Maximal efficiency, when two experiments are performed daily, is reached, apparently, in about 20 days.

(*d*) Practise affects the left hand no more than the right; consequently the index of right-handedness is unaffected by repetition of the test.

(*e*) Practise particularly increases the rate in the later trials, *i. e.,* it particularly affects warming-up, yet "the true practise gain is one mainly in the initial efficiency of performance, as distinguished from the warming-up gain, which shows itself chiefly in continued efficiency of performance" (Wells).

(*f*) An intermission of 10 to 14 days has no unfavorable effect upon practise gains, save that the feeling of fatigue may appear when work is resumed.

(11) *Diurnal rhythm.* Dresslar (10) found evidence of a diurnal rhythm, with a minimum at 8 A. M. and a maximum at 4 P. M. Marsh (18) also found that afternoon records generally surpassed those of the morning, though his figures do not accord very closely with those of Dresslar. Marsh also discovered that the later periods in the evening, which were not tested by Dresslar, furnished the most rapid rates of all. Similarly, Hollingworth noted a tendency toward better performance at the end of the day.

(12) *Dependence on 'general condition.'* When general well-being was ranked as good, medium, below medium and poor, Wells was unable to discern any relation between these several conditions and tapping efficiency, while there was, in the case of susceptibility to fatigue, a tendency, if anything, toward an inverse relation, *i. e.,* fatigue seemed to be greatest on 'good' days. Dresslar's observation that a vigorous walk decreases, while mental work increases speed of tapping has been generally confirmed by other investigators.

In investigating the effect of dental hygiene upon 5th-grade Cincinnati pupils Miss Kohnky found that the children who had special treatment exhibited a decided lowering of the fatigue index (1st 15 compared with last 15 sec.).

The effect upon rate of tapping induced by caffein in various amounts is summarized thus by Hollingworth: "The typical caffein effect on a motor process such as that involved in the tapping test seems to be a stimulation, which is sometimes preceded by a brief and slight initial retardation. The magnitude of the stimulation (*a*) varies directly with the size of the dose, and (*b*) is relatively slight when the caffein is taken in the forenoon." "There is no secondary or after-effect shown within the 72 hours over which the intensive doses were traced."

(13) *Correlation with mental ability and social status.* (*a*) The correlation between tapping ability and mental ability is found to be generally positive by Smedley, Gilbert, Bolton, Kirkpatrick, Burt ($r = .41$ to $.65$) and Abelson ($r = .28$ to $.42$), to be indifferent by Bagley (also by the author), while Binet and Vaschide report a positive correlation with 12-year-old pupils and an inverse correlation with 16 to 20-year-old pupils. While Gilbert found a very marked positive superiority of the 'bright' children in general, the relation did not appear at ages 16 and 17. Bolton found that "good children" (apparently meaning those drawn from the better social classes) were uniformly superior in tapping to children of the poorer class, both with the right and with the left hand. The fact that the divergence is greater at 9 than at 8 years he attributes to a general arrest of development in the poor-class children.

(*b*) Bolton also states that the "good" children showed a distinctly greater practise-improvement—a discovery which he terms "new and significant," and which he thinks is indicative of a fundamental difference in the ability of these two classes of children to take on new habits and profit quickly by experience.

(14) *Coefficient of reliability.* By repeating the test its reliability can be determined. The figures available, those of Burt and of Abelson, refer to a less desirable form of the test, and are conflicting. Burt found the coefficient of reliability to

be only .51, whereas Abelson obtained .92 for a group of girls and .91 for a group of boys. Burt reports that trials with a better form of procedure have yielded distinctly higher coefficients.

(15) *Miscellaneous correlations.* Burt found the following correlations with speed of tapping: with discrimination of pitch .48, with comparison of length of lines .36, with discrimination of lifted weights .42, with esthesiometric limen .10, with card dealing .57, with card sorting .57, with immediate memory .40, with the mirror test .74, with the spot-pattern test .57, with McDougall's dot test .78. Abelson's work with 88 girls and 43 boys in London 'special classes' afforded a somewhat similar series of correlations with eight different tests, though the results for the boys and for the girls were not always accordant. For further details the reader may consult the original article.

(16) *Epilepsy and insanity.* Smith reports inability to discern any characteristic differences between the speed of tapping in epileptic and in normal individuals. In cases of insanity that exhibit retardation, however, Wells finds three characteristic phenomena: (*a*) a reduction in the absolute rate; (*b*) 'reversal,' *i. e.,* acceleration of speed within a series, and (*c*) 'transference,' *i. e.,* superiority of 'following' over the 'preceding' hand. Strong found somewhat the same results, though he concludes that transference is not always seen in depressive states. In the manic states he finds the rate of tapping usually increased over that of the same *S* when normal. Inter-serial warming-up is also characteristic of manic *S*'s.

(15) *Dependence on the type of movement.* The restriction of the tapping movement to specific joints, as has been attempted by some investigators, is difficult to accomplish in practise. However, it appears that the fastest rate is made when the movement is performed by the elbow joint, which is the one mainly concerned in the type of free movement here described. Kelly, for instance, found that the speed of tapping was faster with the forearm than with the forefinger in about the ratio of 15 to 13. From this, in connection with other tests of dexterity, especially tests of minimal movement, he argues

that children only gradually acquire dexterity and quickness of movement with the fingers, and that this passage "from fundamental to accessory," to use Burk's phrase, indicates the necessity of a general readjustment of the motor tasks required of children.

REFERENCES

(1) A. R. Abelson, (*a*) The measurement of mental ability of 'backward' children. *BrJPs*, 4: 1911, 268-314; (*b*) Tests for mental deficiency in childhood. *The Child*, 3: 1912, 1-17.

(2) W. C. Bagley, On the correlation of mental and motor ability in school children. *AmJPs*, 12: 1901, 193-205.

(3) A. Binet and J. Courtier, Recherches graphiques sur la musique. *AnPs*, 2: 1895 (1896), 201-222.

(4) A. Binet and N. Vaschide, (*a*) Epreuves de vitesse chez les jeunes garçons. *AnPs*, 4: 1897 (1898), 64-98. (*b*) Expériences de vitesse chez les jeunes gens. *Ibid.*, 200-224.

(5) C. B. Bliss, Investigations in reaction-time and attention. *SdYalePsLab*, 1: 1893, 1-55.

(6) T. L. Bolton, The relation of motor power to intelligence. *AmJPs*, 14: 1903, 615-631.

(7) W. L. Bryan, On the development of voluntary motor ability. *AmJPs*, 5: 1892, 123-204.

(8) C. Burt, Experimental tests of general intelligence. *BrJPs*, 3: 1909, 94-177.

(9) C. Burt and R. C. Moore, The mental differences between the sexes, *JEPd*, 1: 1912, 273-284, 355-388.

(10) F. B. Dresslar, Some influences which affect the rapidity of voluntary movements. *AmJPs*, 4: 1892, 514-527.

(11) K. Dunlap, New uses for the Ewald chronoscope. *JEdPs*, 4: 1913, 99-101.

(12) J. A. Gilbert, Researches on the mental and physical development of school children. *SdYalePsLab*, 2: 1894, 40-100.

(13) J. A. Gilbert, Researches upon school children and college students. *UnIowaSdPs*, 1: 1897, 1-39.

(14) H. L. Hollingworth, The influence of caffein on mental and motor efficiency. *ArPs* (*e*), No. 22, 1912, N. Y. Especially ch. iv., pp. 25-43.

(15) R. L. Kelly, Psychophysical tests of normal and abnormal children; a comparative study. *PsR*, 10: 1903, 345-372.

(16) E. A. Kirkpatrick, Individual tests of school children. *PsR*, 7: 1900, 274-280.

(17) Emma Kohnky, Paper on effect of dental hygiene on physical and mental efficiency of 5th-grade children, to appear in *JEdPs*.

(18) H. D. Marsh, The diurnal course of efficiency, Columbia Univ. diss., N. Y., 1906. Pp. 99.

(19) J. M. Moore, Studies of fatigue. *SdYalePsLab*, 3: 1895, 68-95.

(20) O. Raif, Ueber Fingerfertigkeit beim Clavierspiel. *ZPs*, 24: 1900, 352.

(21) F. Smedley, *Rep. dep. child-study and pedagogic investigation*. No. 3, 1900-1901 (Chicago Public Schools). Also in *RepComEd*, 1902, 1, 1115-1138.

(22) W. G. Smith, A comparison of some mental and physical tests in their application to epileptics and to normal subjects. *BrJPs*, 1: 1905, 240-260.
(23) E. K. Strong, A comparison between experimental data and clinical results in manic-depressive insanity. *AmJPs*, 24: 1913, 66-98.
(24) Helen B. Thompson, The mental traits of sex. Chicago, 1903. Pp. 188.
(25) F. L. Wells, A neglected measure of fatigue. *AmJPs*, 19: 1908, 345-358.
(26) F. L. Wells, Normal performances in the tapping test before and during practice, with special reference to fatigue phenomena. *AmJPs*, 19: 1908, 437-483.
(27) F. L. Wells, Studies in retardation as given in the fatigue phenomena of the tapping test. *AmJPs*, 20: 1909, 38-59.
(28) F. L. Wells, Sex differences in the tapping test: an interpretation. *AmJPs*, 20: 1909, 353-363.
(29) F. L. Wells, Motor retardation as a manic-depressive symptom. *AmJIns*, 66: July, 1909. (Reprint, pp. 52.)

TEST 11

Accuracy or precision of movement: Aiming.—The purposes for which tests of accuracy of movement have been employed are practically the same as those cited for the tapping test, viz.: to obtain an index of general voluntary motor ability with which to compare different children, to compare the right with the left hand, to determine the development of motor control with age, its differentiation with sex, and to test its correlation with mental ability. These tests have been only rarely used for determining the presence of fatigue, though they have been proposed as means for the diagnosis of incipient ataxia.

Tests of accuracy vary greatly in form: in fact, they virtually shade by degrees from those which prescribe a rapid, accurate movement similar to the tapping movement of Test 10, *e. g.*, the Columbia test described by Wissler (5), to those which prescribe a slow, steady movement more akin to a test for steadiness (No. 13).

Two types of precision test have been selected for consideration: the aiming, or 'target' test, and the line-drawing, or tracing' test (No. 12).

The common feature of all forms of aiming test is the measurement of the extent of error made by an individual when he tries in a series of discrete voluntary movements of hand or arm to hit some form of mark or target. According to the par-

ticular form of movement employed, the test has been known as the 'probing test,' the 'target test,' the 'thrusting test,' etc. These movements have ranged, to speak more specifically, from a simple vertical probing movement of 6 mm. extent (Bryan, 2) to whole-arm aiming with a pencil at a paper target at arm's length (Thompson, 3; Whipple, 4), or making lunging thrusts with a wand, somewhat after the fashion of a fencer, or even throwing ordinary marbles at a bull's-eye target 2 m. distant (Bagley, 1).

The results of any such precision test will obviously be conditioned by the position of the target with respect to *S*, by the extent and rate of the aiming movement, and likewise, though the fact seems not always to have been recognized, by the individually variable improvement in accuracy which will appear if a series of 'shots' are taken at the same target. Hence, to be satisfactory, an aiming test should prescribe and standardize all these conditions; it should also admit of an exact evaluation of each aiming movement. The form of test here described was devised by the author several years ago to meet these conditions, and has proved satisfactory in use. Though the error of a single stroke is large (as is certain to be the case in any form of aiming test), the average of the 30 thrusts made by the same *S* is very constant. The use of ten marks in place of a single mark, or bull's-eye, removes to a large extent the improvement error just mentioned.

Hollingworth, in his investigation of the effect of caffein (see reference, Test 10) used a form of test known as the 'three-hole test,' which demanded both speed and accuracy. The *S*'s were required to thrust a stylus as rapidly as possible into three holes arranged in a triangle, and continue the process until 100 contacts had been registered with the metallic insets in these holes. Here the introduction of the speed factor makes the test different in principle from mere aiming. The same thing is true of the 'irregular dotting' test, developed by MacDougall, and employed by him and by Burt and others in England. In it *S* is required to hit with a pencil an irregularly arranged series of small circles which are moved by clockwork past a narrow slit. The moving tape which carries the circles is speeded up until *S* can just barely succeed by an effort in hitting the marks. This test, when carried out for a prolonged period, may be regarded more as a test of maintenance of attention than of motor coördination.

APPARATUS.—Prepared blanks containing ten crosses irregularly arranged (Fig. 30). A base-board upon which the blanks are fastened, arranged to be secured upon the wall and adjusted to varying heights (Fig. 31). Metronome (Fig. 21). Pencil with tough and moderately hard lead.[1] Millimeter scale.

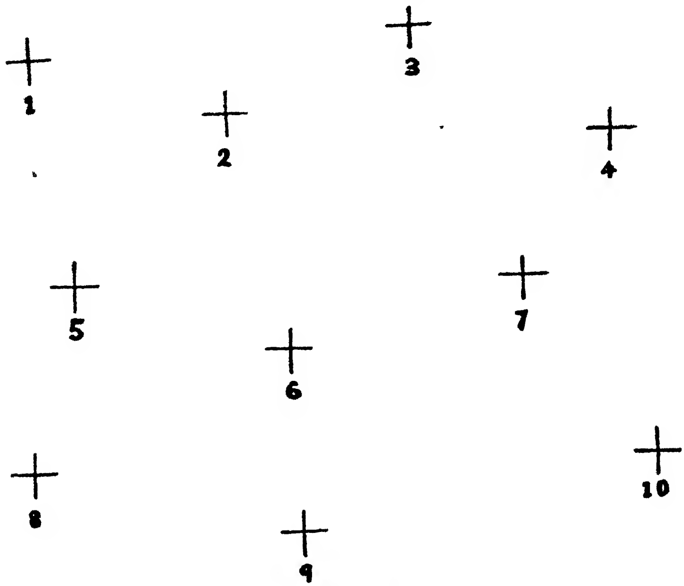


FIG. 30. TARGET BLANK.

The numbers, which are added here to show the order in which the crosses are to be struck in the first round, do not appear on the blank in use. Cut half size.

PRELIMINARIES.—Fasten the board upon the wall and arrange the counterweight properly, so that the board will remain in any position from one to two meters from the floor and will not be displaced when struck by the pencil.

Set the metronome at 138. Fasten the target-sheet upon the board, with the name-date corner in the lower right-hand corner of the board. Place a demonstration target on the wall conveniently near.

METHOD.—Make clear to *S* the following directions: (1) He

[1] By covering the base-board with a felt pad a pointed metal stylus may be substituted for the pencil.

is to stand with his right shoulder (for the right-hand test) squarely in front of the target, at such a distance that his pencil just strikes the target when his arm is fully extended.[1] (2) He is to strike in time with the beat of the metronome, following a two-beat rhythm, so that the pencil hits the target at the one beat and is drawn back at the next. (3) Each stroke is to

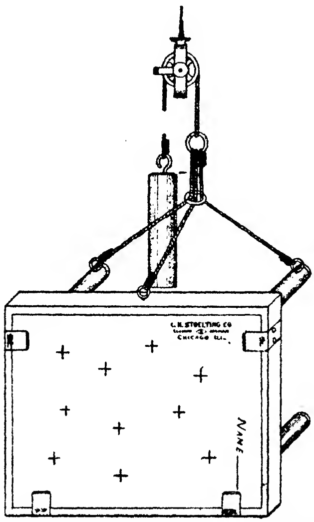


FIG. 31. ADJUSTABLE BASE-BOARD FOR TARGET TEST.

be a full, smooth stroke, not jerky or too short, and the pencil must, therefore, be brought back each time until it touches the shoulder. (4) He is to start at the first cross and make successive strokes, one at each cross in the series until the tenth is reached (see Fig. 30). This process is twice repeated, but in the second round, further to avoid practise, the order is from ten to one. *S* thus makes three shots at each mark, or thirty in all.

Before conducting the test proper, let *S* try the experiment upon the demonstration target. It will save time if *E* also illustrates the process at this time. *E* may assist *S* in follow-

[1] If the subjects are of approximately the same size, this distance may be marked upon the floor by a chalk line.

ing the rhythm of the metronome by saying: 'down, back,' 'down, back,' etc.

Place a fresh sheet upon the target-board and test the left hand.

TREATMENT OF RESULTS.—Measure the error of each thrust by the application of the millimeter scale. A pair of dividers may be helpful in this process. Average the thirty errors and compute the mean variation or standard deviation. Any 'shots' that have struck the lines of the crosses and are difficult to detect may be easily located by reversing the sheet.

RESULTS.—(1) On the basis of similar tests, other investigators have shown a gradual increase in accuracy with *age,* particularly during the years 5 to 8.

(2) *Sex* differences are slight, but, on the whole, boys are more accurate than girls, and men than women.

(3) The author, in using the test as described, has found an error of 4 to 6 mm. in university students, while in a group of fifty 8th-grade boys the following results were obtained:

Right hand, average 5.12, lowest 3.75, highest 8.34.

Left hand, average 6.39, lowest 4.15, highest 9.27.

(4) For the boys just mentioned the correlation between right and left-hand efficiency was 0.54.

REFERENCES

(1) W. C. Bagley, On the correlation of mental and motor ability in school children. *AmJPs,* 12: 1901, 193-205.

(2) W. L. Bryan, On the development of voluntary motor ability. *AmJPs,* 5: 1892, 123-204.

(3) Helen B. Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(4) G. M. Whipple, The influence of forced respiration on psychical and physical activity. *AmJPs,* 9: 1898, 560-571.

(5) C. Wissler, The correlation of mental and physical tests. *PsMon,* 3: No. 6, 1901. Pp. 62.

TEST 12

Accuracy, precision, or steadiness of movement: Tracing.— The purposes for which tracing has been used are identical with those outlined for the preceding test, but the present test differs from the former in that the movement is continuous, analogous

to that made in drawing a line—the so-called 'writing movement' of Bryan (3) or 'tracing test' of Bagley (1). Since steadiness of movement is quite as much in demand as accuracy (of the sort required in Test 11), this test is often classed as a steadiness test, rather than as an accuracy test, but it differs from steadiness tests proper in that it measures control of a voluntary movement, whereas the latter measure the extent of involuntary movement which takes place when the hand or arm is held at rest (Test 13).

The technique most commonly adopted for the tracing or line-drawing test consists in passing a metallic needle or stylus along a narrow slit between metallic strips and noting by telegraph sounder, bell, electric counter, or graphic record, the number of contacts made in passing along a given portion of the slit. This slit may be straight and bounded by parallel sides (Bolton, 2), or by slightly converging strips (Bryan, 3; Thompson, 4), or the slit may in portions be curved, as in the scrolls used by Bagley. Some tests have been made at Columbia University with an irregular printed pattern, which is to be traced by the subject with a lead pencil. In any of these tests the movement may be arranged to involve primarily either the finer muscles of the hand and fingers or the larger muscles concerned in the whole-arm movement.

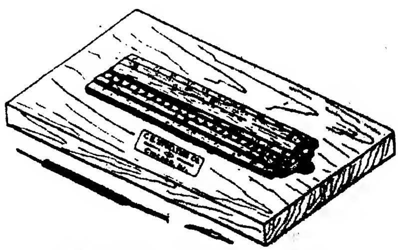


FIG. 32. TRACING-BOARD.

The test here described follows the method used by Bryan and by Thompson.

APPARATUS.—Tracing-board (Fig. 32). Metallic stylus with flexible connecting wire. Telegraph sounder (Fig. 33). Two dry or other open-circuit cells. No. 18 annunciator wire.

PRELIMINARIES.—Wire the battery in series with the tracing-board, the sounder and the stylus.

METHOD.—(1) Seat *S* comfortably with the tracing-board squarely before him and the apex of the angle pointing toward him, so that the movement is directly toward the body in the median plane. Let *S* hold the stylus as he chooses, place the

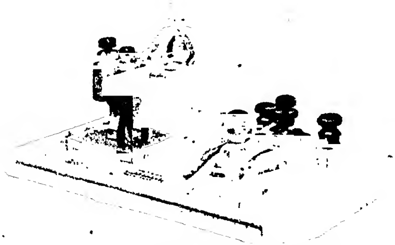


FIG. 33. TELEGRAPH SOUNDER AND KEY.

The sounder may be purchased separately, but the key will be found useful for other experimental work, if only for opening and closing the circuit at will. When provided with a special pointer, this sounder is used for graphic records, as in Test 13.

tip at the opening of the strips, and then attempt to draw a line on the glass between the strips of metal without touching either one. The movement should be *continuous* from start to finish, made entirely free-arm (not a finger or wrist movement, and without supporting the hand or arm in any way). The rate of movement must be illustrated as accurately as possible by *E*, and should be such that the full length of the strips is traversed in 9 sec. Allow *S* two or three preliminary trials, and endeavor to secure approximately this rate before starting the records. As soon as a click of the sounder indicates a contact, *S* is to stop and begin again with the left hand. Repeat until *S* has made five trials with each hand, alternately. *E* records in each case the point on the scale at which contact is made.

(2) Turn the apparatus around 180°, and in the same manner test movement away from the body.

(3) Test movement from left to right and (4) from right to left with either hand by placing the test-board so that the strips lie parallel with the edge of the table nearest *S*.

Variations of Method.—(1) If it is desired not to compare movements in different directions, but merely to compare the right and left-hand efficiency of different *S*'s, the test may with advantage be shortened by adopting the method used by Bryan in his test of school children; viz.: set the test-board so that the strips make an angle of approximately 45° with the edge of the table, *i. e.,* with the right-hand end of the strips turned 45° away from the body for the right-hand test, and with the left-hand end of the strips turned away to the same amount for the left-hand test. Make five tests with movement inward and five with movement outward with each hand.

The conditions may be still further varied (2) by requiring *S* to stand and to hold the stylus at arm's length, (3) by allowing *S* to sit and to support his arm at the elbow, or (4) to support his hand on the base-board while executing a forearm or whole-arm movement. The last was the method followed by Miss Thompson.

Treatment of Results.—The simplest treatment of the data is to secure an index of precision by averaging the distances at which the several points of contact are made. For a more complex method of computing the measure of precision the reader is referred to Bryan, pp. 180 ff.

Results.—(1) There is greater *variation* in the outcome of precision or steadiness of movement tests than in that of rate of movement tests.

(2) There is undoubtedly a more or less constant improvement in precision with *age,* but sufficient data are not yet at hand to determine the yearly increments clearly. There is certainly, however, a decided gain during the years 6 to 8, while Bolton also noted improvement from the 8th to the 9th year of age.

(3) *Dextrality.* In general, the right hand is, of course, distinctly superior to the left, but the amount of this superiority varies remarkably with age, and is, according to Bryan, less evident at 15 and 16 than at 6, 9 or 12 years of age.

Right-handed adults sometimes make better records with the left hand, as a rule because they perform the test relatively carelessly with the right and cautiously with the left hand.

(4) *Sex.* With either hand, boys are probably slightly superior to girls, for, while Bolton reports the superiority of the girls in his groups of children, Bryan's examination of some 700 children revealed the following relations when the results for both hands were computed together: boys were superior to girls in 51.5 per cent. of the trials, girls superior to boys in 35.3 per cent., and the sexes equal in 13.4 per cent. Moreover, in a similar test, Thompson found men superior to women.

(5) Both Bolton and Thompson found *movements inward* or toward the body uniformly steadier than movements outward or away from the body.

(6) *Correlations.* Bagley found "a decidedly inverse relation between mental ability, as indicated by class standings and motor ability, as indicated by the tracing test." But Bolton, who used a different form of test, and apparently correlated rather with the social status than with the class standing, reports that 'good' are steadier than 'poor' children.

REFERENCES

(1) W. C. Bagley, On the correlation of mental and motor ability in school children. *AmJPs*, 12: 1901, 193-205.

(2) T. L. Bolton, The relation of motor power to intelligence. *AmJPs*, 14: 1903, 615-631.

(3) W. L. Bryan, On the development of voluntary motor ability. *AmJPs*, 5: 1892, 123-204.

(4) Helen B. Thompson, The mental traits of sex. Chicago, 1903. Pp. 188.

TEST 13

Steadiness of motor control: Involuntary movement. — The general idea in this type of test is to measure the amount of involuntary movement which appears when the finger, hand, arm or body as a whole is held as nearly motionless as possible. Like the preceding tests, this has been frequently employed as a means of obtaining an index of motor ability (Bagley, 1; Bolton, 2; Hancock, 6), but it has also been employed for numerous other purposes, *e. g.*, for examining the motor tendencies accompanying ideational activity (Jastrow, 7, 8; Tucker, 15), for examining the bodily expression of affective states (Titchener, 14), for examining the nature of constant tendencies

toward automatic movements, or the possibility of developing such movements by training (Thompson, 13; Solomons and Stein, 10; Stein, 12), and for detecting the presence of incipient or recent chorea (Crichton-Browne, 4).

This variety of purpose illustrates very forcibly the difficulty of so conducting the test as to examine any one phase alone—a difficulty which is aggravated by the fact that the tracings of involuntary movement are often affected, not only by the factors implied above, but also to a considerable extent by the direction of the attention, by the relative position of the body and the instrument, and by physiological processes, especially respiration. It is, therefore, not surprising that many writers, *e. g.,* Bagley and Bolton, have incorporated records of involuntary movement only with qualifications and without placing much insistence on their worth.

The instruments most commonly employed are the ataxiagraph (described by Crichton-Browne, used by him and by Hancock, Wallin, Bolton and others for measuring the swaying of the body as a whole), the tremograph (described by Bullard and Brackett, 3, and also used by Hancock) for testing the arm or finger, the automatograph of Jastrow or that of Stein, further elaborated by Titchener, for measuring involuntary movements of the arm; the digitalgraph devised by Delabarre (5), described, together with others of the instruments just named, by MacDonald (9), and used for recording the tremor of the finger, and several instruments, as yet unnamed, for testing either the arm or hand, according to the conditions of their use.[1]

Apparatus.—Brass plate, set at an angle of 45°, and pierced

[1]For some purposes it is probably quite as well to test the subject without the use of apparatus. Thus, according to Sturgis, the following constitutes an infallible test for chorea: "Bid the child hold up both hands open, with extended arms, the palms toward you. If that is done steadily, both hands upright and both alike, no finger or hand quivering, no falling back of either hand, nothing to choose between the positions of the two, then the child has not, nor is it near (either before or after) St. Vitus' dance. You may confirm this test by another. Let the child place the open hands upon yours, palm to palm. Look then at the backs of the child's hands, observe whether fingers or thumbs (especially the latter) repose without tremor and without restraint."

with a series of holes whose diameters are 32, 20, 16, 13, 11, 10, 9, 8 and 7 sixty-fourths of an inch, respectively (Fig. 34). Metallic needle of special design, with flexible connecting wire.

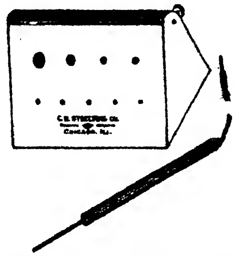


FIG. 34. STEADINESS TESTER.

Telegraph sounder (Fig. 33), with writing lever attached to the armature, and with sending key (or separate short-circuiting key). Kymograph (Fig. 22), with accessories. Stop-watch (Fig 20). Four dry or other open-circuit cells. Insulated connecting wire. [A low table, about 65 cm. high, and an adjustable chair of the typewriter chair pattern are desirable.]

PRELIMINARIES.—Wire the needle, brass plate, sounder and key in series with the battery in such a way that contact between the needle and the plate will actuate the sounder. Place smoked paper on the kymograph drum, and adjust to approximately one revolution in 40 sec. Adjust the position of the sounder so that the writing lever gives a satisfactory tracing on the drum.[1]

METHOD.—Seat *S* before the table in a comfortable position. Place the brass plate flush with the front edge of the table, in front of *S*'s right shoulder for the right-hand test, and in front of his left shoulder for the left-hand test. Instruct *S* to hold the needle in such a way that his finger tips are in contact with the expanded flange of the holder, and, at command, to hold the tip of the needle within the largest hole, and to maintain this position, so far as possible, without touching the brass plate during the 15 sec. allowed for the trial. *S*'s hand and arm must be entirely free from all support or contact with his

[1] For an account of the manipulation of the kymograph, consult Test 10.

body or other object, and his forearm should form an angle of approximately 100° with his upper arm. The needle should be inserted about 6 mm. into the hole.

Show *S* that the click of the sounder will serve as warning for him that the needle is making contact with the plate.

In conducting the test, allow *S* about 3 sec. for taking the position (since a certain amount of movement will appear when the needle is first inserted that will afterward be checked by *S*'s control), then close the short-circuiting key and at the same time start the stop-watch. At the expiration of 15 sec. open the circuit, stop the watch and the kymograph, and at once rearrange the instruments for the left-hand test.

Allow 30 sec. rest, and then test the right hand and the left hand with the next smaller hole, and so on, until a hole is reached so small that *S* has reached the limit of his capacity and is clearly unable to keep the needle free from contact with the plate.

In most cases a few unrecorded preliminary trials will show approximately the degree of *S*'s control, and the tests with the larger holes may be omitted with a consequent saving in time.

Variations of Method.—The test may be modified by altering the relative position of the plate and *S*'s body as follows: (*a*) by requiring *S* to stand and to hold the needle extended at arm's length, (*b*) by allowing *S* to rest his elbow upon the table with the forearm free, or (*c*) by supporting his forearm and wrist and testing the steadiness of the hand and fingers, and (*d*) by extending the time to 30 or 60 sec. The results will naturally differ characteristically from those obtained under the conditions prescribed as standard.

Treatment of Results. — By reference to the kymograph tracings, count the number of contacts for each hole. For comparative purposes, *E* may take the total number of contacts made in a given series of holes or the number made in that hole which most satisfactorily tests the steadiness of the subjects under investigation.

RESULTS.[1]—(1) In all tests of involuntary movement it is clearly seen that *age* is an important factor. Hancock concludes that adults have approximately 5.8 times as much control over their fingers as do children aged 5 to 7 years.

(2) Distinct *sex* differences have not been established.

(3) As measured on the ataxiagraph, involuntary swaying was found by Wallin to be clearly less for the moron than for the imbecile group of epileptics, and slightly less for males than for females.

NOTES.—In rare cases the use of the graphic method may be dispensed with, especially if *E* proves by practise very skillful in the correct counting of the rapid and irregular strokes made by the sounder when the test approaches the limit of *S*'s capacity, but this simplification is not recommended, because it is exceedingly difficult to make the count under these conditions (cf. Bolton), and, moreover, there may appear very short, rapid contacts that will actuate the sounder sufficiently to produce a noticeable indication on the tracing, but not sufficiently to produce a noticeable click. Again, *S* will occasionally make contacts of long duration. With the graphic record it is possible for *E* to measure the duration of these contacts, and, if desired, to base *S*'s record upon the proportion of the time during the 15 sec. that is occupied in contact, rather than upon the number of contacts alone.

The ordinary electric counter is not recommended as a substitute for the graphic method, because, as mentioned in Test 10, it will not operate reliably when the contact is very brief, even although a large number of cells are used in the battery. To avoid the possibility of a similar error in the use of the sounder, the excursion of the armature should be rather short, *i. e.*, the armature should be adjusted as near the fields as is possible if it is to give a clean stroke.

If involuntary tremor is to be studied with special care, *e. g.*,

[1] A summary of the results obtained with the automatograph is given by Jastrow (8, pp. 307-336.)

if *E* wishes to make an extended study of an individual case, a more sensitive instrument should be employed. For such work the tridimensional analyzer of Sommer (11), also described by Titchener (14), is recommended.

REFERENCES

(1) W. C. Bagley, On the correlation of mental and motor ability in school children. *AmJPs*, 12: 1901, 193-205.

(2) T. L. Bolton, The relation of motor power to intelligence. *AmJPs*, 14: 1903, 615-631.

(3) Bullard and Brackett, *Boston Med. and Surg. Jour.*, 2: 1888, 600.

(4) J. Crichton-Browne, The nervous system and education, being part 4 in The book of health, edited by M. Morris, London, 1883 (now out of print).

(5) E. B. Delabarre, Ueber Bewegungsempfindungen, 1891. See also *AnPs*, 1: 1894 (1895), 532.

(6) J. A. Hancock, A preliminary study of motor ability. *PdSe*, 3: 1894, 9-29.

(7) J. Jastrow, (*a*) A study of involuntary movements. *AmJPs*, 4: 1892, 398-407. (*b*) A further study of involuntary movement. *Ibid.*, 5: 1892, 223-231.

(8) J. Jastrow, Fact and fable in psychology. Boston, 1900. Pp. 370.

(9) A. MacDonald, Experimental study of school children, etc. Chs. 21 and 25, *RepComEd*, 1897-8 (1899).

(10) L. Solomons and G. Stein, Normal motor automatisms. *PsR*, 3: 1896, 492-512.

(11) R. Sommer, Lehrbuch der psychopathologischen Untersuchungsmethoden. Berlin, 1899. Pp. 399.

(12) Gertrude Stein, Cultivated motor automatisms. *PsR*, 5: 1898, 295-306.

(13) Helen Thompson, The mental traits of sex. Chicago, 1903. Pp. 188.

(14) E. B. Titchener, Experimental psychology. 2 vols., N. Y., 1901 and 1905.

(15) M. A. Tucker, Comparative observations on the involuntary movements of adults and children. *AmJPs*, 8: 1897, 394-404.

(16) J. E. Wallin, Experimental studies of mental defectives. *EdPs Mon*, No. 7, Baltimore, 1912.

CHAPTER VI

TESTS OF SENSORY CAPACITY

Psychophysical tests of sensory capacity are divisible into tests of liminal sensitivity (sensitivity proper) and tests of discriminative or differential sensitivity (sensible discrimination.)[1] In the former, we measure the bare capacity of experiencing sensations, the minimally perceptible stimulus or *stimulus limen, e. g.,* the lightest pressure that can be felt, the least intensity of tone that can be heard, etc.; in the latter, we experience different sensations and report upon their difference; we seek, in other words, to determine the minimal objective difference of stimulation that can just be mentally cognized as different, to determine the *difference limen, e. g.,* the smallest change of vibration-rate that will suffice to yield two perceptibly different tones, or the smallest difference in weight that can just be recognized as a difference.

These two measurements of sensory capacity, liminal and discriminative sensitivity, can be applied to any modality, *i. e.,* to any sense-department, and to any attribute of sensation, *i. e.,* to quality, intensity, extent, and duration. We may measure, for instance, in the case of the ear, liminal sensitivity in terms of intensity, discriminative sensitivity in terms of intensity, in terms of quality, in terms of duration, etc. We may, furthermore, determine the total number of auditory sensations that can be experienced, *i. e.,* modal sensitivity.

In the case of the two sense-departments that possess the attribute of extent, visual and cutaneous sensations, we may also measure the capacity to discriminate difference in the localization of two stimuli, or the limen for spatial discrimination. While, strictly speaking, this may be regarded as a more complex process than simple sensory discrimination, it may, for our purposes, be included as a sensory test. Indeed, with

[1]For a discussion of the terminology, methods, purposes, and results of psychophysical methods, consult Titchener, *Experimental Psychology*, especially, Vol. 2, part 2.

both the eye and the skin, this determination, for practical purposes of exploration of the sense-organ or of the measurement of its functional capacity, has superseded the determination of liminal sensitivity. Thus, in visual sensation, visual acuity refers, not to the liminal sensitivity of the retina for stimulation, but, in principle, to the capacity to distinguish the separation of two points. Similarly, in cutaneous sensation, for a practical test of functional capacity, the determination of the 'limen of duality,' as it may be termed, by means of the esthesiometer, has been more often employed than the simple determination of liminal sensitivity to pressure.

Now, the quantitative determination of sensitivity in the psychological laboratory has given rise to a most elaborate and refined methodology, and has, in fact, been the chief problem of the science of psychophysics. It is not the purpose here to discuss or duplicate these exact methods, but merely to indicate the manner in which, for comparative purposes, one may secure an index of functional efficiency by empirical methods. It must be clearly understood that the determination of an exact stimulus or difference limen in the psychological laboratory, with minute introspective analysis of the factors that condition the process and with elaborate methodological procedure, is quite a different process from this simple determination of functional capacity for comparative purposes. If, for convenience, the technical terms of psychophysics are here employed, they are employed with this qualification in mind.

To make this point clearer: the procedure for the determination of discriminative capacity as herein recommended is not identical with any of the established psychophysical methods. *E* allows *S* at first to try various stimulus differences ranging from large to small, until *S* has acquired general familiarity with the test, and *E* has obtained a general notion of his capacity. *E* may next test *S* more formally by applying a series of stimulus differences ranging from clear subjective difference to subjective equality. He then selects a difference which seems likely to be just cognizable by the subject and applies this difference ten times, with proper reversals for time or space errors. If eight right judgments are given, he then

corroborates the result by trying similarly a slightly smaller, and finally a slightly larger difference, to see if *S* gives in the former case fewer, and in the latter case more correct judgments. *S* knows that a difference exists, but is ignorant of its spatial or temporal position. We thus obtain an index of capacity, but do not determine the mean difference limen, nor even the lower limen, in the psychophysical sense.

Sensory tests of this empirical sort have been employed, partly in connection with the psychology of individual and sex differences, partly in the objective study of general intelligence, partly in the exploration of sense organs for the determination of their working condition, *i. e.*, for hygienic and diagnostic purposes. In all of these fields the emphasis is upon the examination of simple functional capacity, without particular reference to introspective examination or analysis of the accompanying consciousness.

The use of sensory tests in correlation work is particularly interesting. In general, some writers are convinced that keen discrimination is a prerequisite to keen intelligence, while others are equally convinced that intelligence is essentially conditioned by 'higher' processes, and only remotely by sensory capacity—barring, of course, such diminution of capacity as to interfere seriously with the experiencing of sensations, as in partial deafness or partial loss of vision.

While it is scarcely the place here to discuss the evolutionary significance of discriminative sensitivity, it may be pointed out that the normal capacity is many times in excess of the actual demands of life, and that it is consequently difficult to understand why nature has been so prolific and generous; to understand, in other words, what is the sanction for the seemingly hypertrophied discriminative capacity of the human sense organs. The usual 'teleological explanations' of our sensory life fail to account for this discrepancy. Again, the very fact of the existence of this surplus capacity seems to negative at the outset the notion that sensory capacity can be a conditioning factor in intelligence—with the qualification already noted.

The tests which follow are selected from a large number of theoretically possible tests, because of their prominence in such experimental studies as have been mentioned. Their classification is simply by sense-departments. Tests for the exploration of the organ, measurement of its defects, determination of acuity, liminal sensitivity, discriminative sensitivity, in so far as they are described, are given successively for each of the main sense departments.

TEST 14

Visual acuity.—The functional capacity of the eye is examined primarily, of course, for practical purposes in connection with hygienic investigation. Occasionally, it becomes desirable to determine the presence or absence of visual defect in connection with the administration of some mental test, *e. g.*, the cancellation test.

Visual acuity has been studied in its relation to school standing, general intelligence, occupation, habitat, race, sex, as well as to bodily disturbances, such as headache, chorea, indigestion, or other optical defects, such as strabismus, total color-blindness, etc.

Optical inefficiency, aside from color-blindness, may be due to *amblyopia* (dimness of sight not due to refractive errors or demonstrable lesion), or to *asthenopia* (general impairment of retinal efficiency due to anæmia, over-use, etc., and often yielding to proper medical treatment), but is more commonly some form of *ametropia* (defect in shape of the eye-ball, lens, or cornea, with resultant defect in refraction and in the formation of the retinal image).

Ametropia may exist as presbyopia, myopia, hyperopia, or astigmatism.

Presbyopia is the long-sightedness of old age, due to the lessened elasticity of the lens, and is not commonly present before forty.

Myopia, or short-sightedness, is commonly produced by too long an eyeball, the effect of which is to allow rays of light in distant vision to focus in front of the retina and hence to produce a blurred image when they finally impinge upon the retina. The myopic eye is thus unable by any effort clearly to see objects situated at distances of 2 m. or more away, while its 'near-point,' *i. e.*, the nearest point at which clear vision is possible, is brought correspondingly closer, so that objects may be seen clearly when 5 or 6 cm. distant—something which would be impossible for the normal (emmetropic) eye. Pure myopia, as a rule, causes no eye-strain, but it is nevertheless a serious condition, because of its tendency to increase in degree, and because of the appearance in many cases of concomitant pathological disturbances of the retina, which, in extreme cases, result in actual blindness. In practise, moreover, myopia is rarely found pure, but complicated with astigmatism and other defects. True myopia may be counteracted, and its progress checked, but not cured, by the use of properly fitted concave lenses, supplemented by the exercise of caution in the use of the eyes.

Hyperopia, hypermetropia, or long-sightedness, more exactly over-sightedness, is commonly produced by too short an eye-ball, the effect of which is to intercept rays of light too soon, *i. e.*, before they are brought to a normal focus. The hyperopic eye must consequently exert an effort

of accommodation in order clearly to see objects at a distance, while for near work this effort must be excessive. The result is that the hyperopic eye is under constant and abnormal strain from the incessant demands upon its ciliary muscle, and that, in consequence, numerous secondary symptoms or resultant effects appear, some of them obvious, others unexpected, many of them serious. Local symptoms appear in inflammation, redness, or soreness of the eyes, lids, or conjunctiva, and in twitching and pain within the eye-ball. Aside from these local disturbances, perhaps the most constant symptom of hyperopia is frontal or occipital headache. Characteristic also is the holding of the work at some distance from the eyes,[1] a peering or frowning expression, and dislike of near work. Eye-strain, whether hyperopic or astigmatic, may also occasion more serious physiological disturbances, such as chorea, vomiting, nervous dyspepsia, etc.[2] Since the hyperopic eye can see clearly at a distance and can read (as its possessor often boasts) with the book held at some distance, the defect is often unsuspected, because the secondary symptoms are not correctly interpreted. On this account, too, it becomes necessary to take special steps to detect its presence, and many of the simple distance tests that have been applied wholesale upon school children utterly fail to diagnose it. The oculist commonly makes use of homatropin or some other cycloplegic to paralyze temporarily the ciliary muscle and thus prevent accommodation. Hyperopia may, however, be detected, though less accurately, by the use of suitable test-lenses, as described below. The defect is counteracted by the use of properly fitted convex lenses.

Astigmatism is produced by an uneven radius of curvature, usually of the cornea; this surface, which should normally be approximately spherical in form, is, in astigmatism, more strongly curved in one axis or meridian than in another, so that the cornea is ellipsoidal in form, *e. g.*, like the bowl of a spoon, or the side, rather than the end of an egg. Thus the eye is double-focussed, and it is impossible by any effort to focus an image clearly in both meridians simultaneously. In measuring astigmatism it is evident that one must assign both the degree of refractive error and the axis in which the error lies, and that in correcting it, a cylindrical lens of the proper curvature must be placed before the eye at exactly the proper axis to counteract the indicated deficiency. This lens only counteracts the defect, and does not cure it. Astigmatism may be in part congenital, in part a phenomenon of growth (often attributable to the pressure upon the eye-ball of the eye-lids and contracted brows, with the result that the maximal refractive index lies at or near the vertical meridian). When present in large amount it becomes a serious obstacle to vision; when present in small amounts, as is apt to be the case in many eyes, it is the occasion of the same phenomena of eye-strain that have been mentioned as accessory to hyperopia; astigmatic headache is particularly symptomatic—indeed, 90 per cent. of all headaches are said to be traceable to this source.

It must be understood that these three defects may, and commonly do, appear in combination, particularly astigmatism with hyperopia or

[1] In high grades of hyperopia, distinct images cannot be secured even by this process, so the child may abandon the attempt to secure clearness and seek merely to increase the size of the image by holding his book near his eyes. He may thus be falsely rated as near-sighted by the casual observer.

[2] The injurious effects of eye-strain have found a special expositor in Dr. G. M. Gould.

myopia, and that the defects may be, and commonly are, unlike in the two eyes of the same individual. Partly for this reason, the proper fitting of glasses is an art, and, like any art, requires great skill, complete familiarity with the conditions, and long practical experience. The tests which are here described make no pretence to exactitude, but are designed to determine, in so far as is possible by simple methods, the existence of defects that *should invariably be referred to a specialist for further examination and treatment.*

For the examination of refraction the chief appliances are (1) the ophthalmometer, for the exact measurement of the degree and axis of astigmatism, (2) the ophthalmoscope, for the examination of the retina, (2) the retinoscope and the skiascope, for the objective determination of refractive errors, (4) test-types and trial lenses, for actual visual tests under varying conditions. While retinoscopy is a method of great value, especially in testing young children, the test-type is, in general, the court of final appeal and constitutes the most widely used and perhaps the most valuable single means for testing visual acuity. The most varied kinds of test-type have been devised by oculists. Perhaps oldest and best known are Snellen's "Optotypi," which form the basis of the tests ordinarily used. Interesting variations are seen in Dennett's Monoyer type, Landolt's C-test, Cohn's E-test and McCallie's Vision Tests (literate and illiterate set).[1]

The first test for ametropia, which is described just below, is based upon the recommendations of the American Ophthalmological Society. The second test supplements the first by detecting astigmatism. Tests 15, 16, and 17 may be added.

A. TEST FOR MYOPIA AND HYPEROPIA

APPARATUS.—Lowell's test-type. Trial frame (Fig. 35). Two —.75 D.[2] and two + .75 D. spherical lenses, and one blank disc. Meter stick.

[1]The McCallie cards, purchasable of Edwin Fitzgeorge, Box 67, Trenton, N. J., are well adapted for rapid schoolroom tests, where the method of distance alone is used and no direct attempt is made to detect hyperopia or astigmatism. The illiterate set is especially ingenious.

[2]One diopter, or dioptric, D, is the amount of refraction given by a lens whose principal focal distance is 1 meter, *i. e.*, the radius of curvature of its surface has a length of 1 meter. This unit is now commonly used in place of the older system of indicating strength in inches of focal length.

PRELIMINARIES.—Place the test type on the wall or stand, on a level with S's eyes, in a strong even illumination, though not in actual sun-light.[1] Seat S comfortably at a distance of 6 m. from the chart.

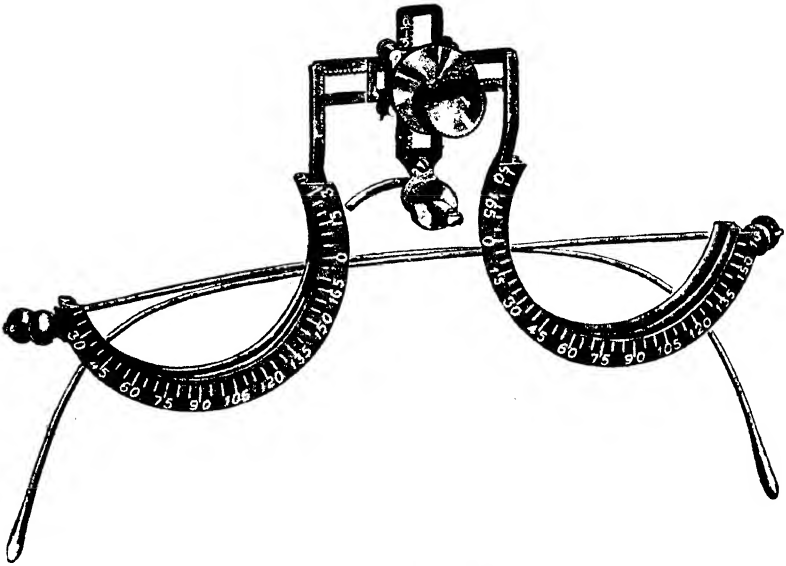


FIG. 35. TRIAL FRAME.

Holds two pairs of lenses, and has rack and pinion adjustment for pupillary distance, and for vertical, and back and forward movement of the nosepiece.

Note any indication of soreness or inflammation of the eyes, lids, or conjunctiva. Ascertain if S has ever suffered from such inflammations, from habitual headaches, or watery eyes; whether his eyes become painful, sore, or strained in doing close work; whether he has previously been examined, and if so, with what result; whether he has ever worn glasses; if they have

[1]If conditions render daylight illumination unreliable or unsatisfactory, E must arrange artificial illumination, carefully shaded from S. Excellent devices for this purpose may be purchased from dealers in optical supplies, e. g., No. 4274, catalog of E. B. Meyrowitz, New York City, price $10. Whatever the source of illumination, the light must not shine in S's eyes. An experienced E may compensate for inadequate illumination by placing S at a distance less than 6 m.—the amount of the correction being determined by E's own acuity under the prevailing conditions.

been worn and discarded, ascertain when and why. If *S* wears glasses, record that fact; test his vision both with and without his glasses, unless time forbids, in which case test with them.

Adjust the pupillary distance and the nose-piece of the trial frame so that the lenses will be centered before *S*'s eyes.

METHOD.—(1) Place the solid disc in the frame before *S*'s left eye. Instruct him always to keep both eyes open. Ask him to read the letters above the red strip[1] (line numbered 8 on the right-hand margin and designated on the chart as standard at 6 m., or 19.7 feet). If this line can be read entire, or all the letters but one, try the next smaller letters, Line 9, but if the standard line cannot be read, try some line above, like 5 or 3. When *S* hesitates, encourage him to guess at the letters. *S*'s performance is recorded most easily by reference to the small figures at the left end of the smallest line that he can read, *e. g.*, if Line 8, vision of the right eye (V. R. E.) = 1; if Line 7, then V. R. E. = 0.8. If *S* cannot read the largest letter, Line 1, we know that V. R. E. = — .1.

(2) *Whatever* the result of this first trial, always next place the + .75 D. lens in the trial frame. This will blur the vision if the eye is emmetropic, so that, if before, V. R. E. = 1., and if vision is now blurred, record V. R. E. = 1. Em.

If *S* can, with the plus lens, read the same line as before, or a smaller line than before, then the eye is hyperopic. Thus, if previously the 10-meter line was read and now the 7.5-meter line, the record will be, V. R. E. = .6 + Hy. = .8, or, if no improvement appeared, V. R. E. = .6 + Hy. = .6.

[1]The red and green strips placed under Lines 8 and 6, respectively, of the Lowell chart are intended to assist E in designating to *S* the lines to be read, and it is assumed that color-blindness may be detected in this manner by characteristic mistakes on the part of *S*. But, as will be understood from what is said later in Test 16, it does not necessarily follow that all color-blind *S*'s would confuse the particular green and red that are placed on this chart. Besides these colored strips, each test-line of the Lowell chart is designated in three ways: the large figures at the right of each line simply number the lines from above downward, Nos. 1 to 11, and serve merely as convenient designations in indicating to *S* the line to be read; just above each test-line there is indicated in meters and in feet the standard distance at which the line should be read; at the left of each line are figures which show at a glance the rating of the line, for measuring visual acuity, in comparison with the 6-meter line (Line 8).

(3) If, in the first test, vision is less than 1.0, and if, in the second test, vision is impaired by the convex lens, then next replace the convex lens by the — .75 D. lens. If vision is now improved so that a smaller line is read, then the eye is myopic, and may be recorded thus; V. R. E. = .6 + My. = .8, or V. R. E. = .6 + My. = 1.[1]

(4) Place the solid disc before the right eye, and test the left eye similarly. Record the results for each eye separately, *e. g.*, V. R. E. = 1. Em. V. L. E. = .6 + My. = .8.

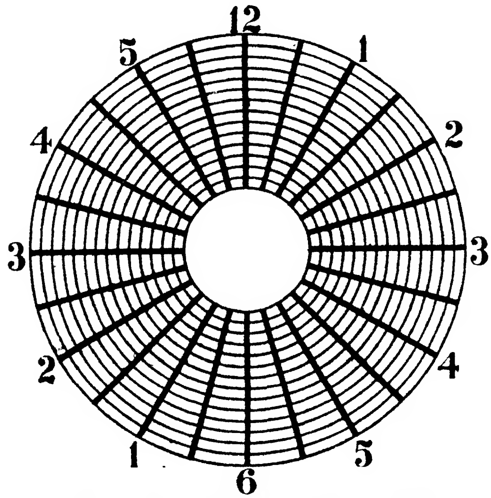


FIG. 36. VERHOEFF'S ASTIGMATIC CHART.
About one-fifth actual size.

B. TEST FOR ASTIGMATISM

APPARATUS.—Trial frame and lenses as above. Verhoeff's astigmatic chart (Fig. 36).[2]

PRELIMINARIES.—Place the chart on the wall, and seat *S* as in the previous test. Be sure that *S*'s head is held squarely erect. If *S* has been found to be myopic or hyperopic, place in the trial frame the lenses which correct, at least partially, this defect.

---

[1]This test may be nullified in some instances by a ciliary spasm in a hyperopic eye which may simulate myopia of almost any degree.

[2]Any standard astigmatic chart may be substituted, but Verhoeff's is, in the author's judgment, best adapted both for making evident the presence of astigmatism and for determining approximately its axis.

Method.—Place the solid disc in the frame before the left eye. Ask *S* whether one or more of the radiating lines seem to him sharper and blacker than those at right angles to them. If he answers in the affirmative, astigmatism is present. This result may be confirmed by causing *S* to move his head from one shoulder to the other, in which case the location of the sharpest lines should shift in a corresponding manner. The amount of astigmatism may be roughly judged by the positiveness and readiness of *S*'s answer; its axis may be determined approximately by his designation of the blackest line or lines.

Place the disc before the right, and test similarly the left eye.

Since astigmatism may exist either alone or in combination with some form of ametropia, it should, when found, be recorded with the previous determination, *e. g.*, V. R. E. = .6 + My. = .8 + As.

If vision is .6 or less, but no form of ametropia can be demonstrated, the defect is recorded as amblyopia, *e. g.*, V. L. E. = .6 + Am.

To summarize the two tests: emmetropia is indicated (unless strain symptoms point to concealed hyperopia) by the reading of the line standard for the distance used and subsequent blurring by the convex lens, hyperopia by improvement or lack of impairment of vision by the plus lens, myopia by vision less than 1., which is improved by the concave lens, astigmatism by unequal clearness of the radiating lines, amblyopia by vision .6 or less without demonstrable refractive error.

Results.[1]—(1) The *frequency* of defective visual acuity is somewhat difficult to state accurately owing to the differences in method and in degree of rigor and precision that have characterized the many investigations upon this point. In especial, a great many investigations in school systems have been made by simple distance tests without the aid of lenses, so that hyperopia, the most frequent defect, has gone practically unmeasured. The general outcome of these simple tests is quite uniform, viz: that one child in three in the public schools suffers

[1]For a general discussion of the examination of eyesight, with special reference to the eyesight of school children, consult Barry, Calhoun, Carter, Cohn, Gould, Hope and Browne, Kotelmann, Newsholme, Risley, Schmidt-Rimpler, Snell, Stilling and Young.

from visual defect. Typical figures are those obtained by Welch at Passaic, and Smedley at Chicago; the latter reports that 32 per cent. of the 2030 boys and 37 per cent. of the 2735 girls examined were defective in vision. While more than half of these defects are of a minor degree, yet, as already indicated, these may be productive of immediate distress and entail serious consequences if neglected.

On the other hand, examinations that have been conducted by skilled opthalmologists with some refinement of method indicate a much larger percentage of defect.[1] Risley's figures indicate that it would be more correct to state that seven children in eight, than that one in three, are ametropic. As chairman of the Philadelphia committee that examined some 2500 children, he gives the refraction at 8½ years of age as hyperopic in 88.11 per cent. of cases, emmetropic in 7.01 per cent., and myopic in 4.27 per cent.: at 17.5 years as hyperopic in 66.84 per cent., emmetropic in 12.28 per cent. and myopic in 19.33 per cent.

(2) From the above figures we may conclude that the *eye in early childhood* is an incomplete eye, naturally underfocussed and poorly adapted for near work. But, as general bodily maturity approaches, the eye under optimal conditions apparently tends to become emmetropic. Conditions of modern life, however, are not optimal for the eye, but rather encourage overuse and neglect, with the consequence that these, at least when added to astigmatism or other congenital defects, tend to develop eye-strain, myopia, and other disturbances of vision.

(3) The relation of overuse and neglect, particularly in the form of excessive near work, to *myopia* is a matter of much dispute. The views currently held by many hygienists are typified by the position of Cohn, who contends that myopia is essentially a disease of civilization and culture; that it is infrequent in peasants and those who lead an outdoor life, and progressively more prevalent and of higher degree as persistent study and near work continue. Thus, in gymnasia, he found that the percentage of myopia increased during six years of study in the following manner: 12.5, 18.2, 23.7, 31.0, 41.3, 55.8. Similarly,

[1]Indeed, there are specialists who assert that an absolutely perfect pair of eyes does not exist.

of 138 pupils at the Friedrich gymnasium who were examined twice at an interval of 18 months, he found at the first test 70 normal and 54 myopic, whereas in the second test, 14 of the 70 had become myopic, 28 of the original myopes had developed a higher degree, and in 10 per cent. serious structural changes had taken place in the retina.

While statistics like these, which indicate a close relation between myopia and near work could be multiplied almost indefinitely,[1] they have not prevented other investigators, chiefly anatomists, physiologists and biologists, from attributing the causation of myopia to heredity rather than to environmental conditions (cf. Barrington and Pearson, Helmholtz). It would seem safe, at least, to conclude that the predisposing causes of myopia are frequently inherited, whatever may be the rôle of poor hygienic conditions in bringing out the defect.

(4) Smedley, on the basis of his method of correlation by grades, asserts that "a smaller per cent. of the *pupils at and above grade* have defective sight than those below grade."

(5) Smedley further demonstrates that defective vision is extremely common in *backward and troublesome children,* and that this fact may be a partial explanation of the behavior of such children. Thus, according to the tests employed at Chicago, 48 per cent of the boys of the John Worthy School were subnormal in visual acuity, as contrasted with 28 per cent of the boys of the same average age in other schools. Moreover, "many of the John Worthy boys had strabismus, hypermetropia, and astigmatism, conditions which would induce asthenopia when the eyes were used in close and long application to books, and it is easy to believe," adds Smedley, "that the strain thus set up when an attempt was made to study was a factor in producing dislike for school and subsequent truancy."

(6) Van Biervliet (22) has sought to obtain a correlation between the visual acuity test and *intelligence,* not by direct

[1]Dürr, for example, in explanation of the alarming prevalence of myopia in Germany as contrasted with other countries, has sought to show its dependence upon the excessive demands of the German school system: he estimates that, during the years 10 to 19, the typical English boy spends in study 16,500 hours, in exercise 4500; the French boy, in study, 19,000, in exercise 1300; and the German boy, in study 20.000, and in exercise 650.

reference to visual acuity itself, but to the mean variation measured by a series of tests of acuity, *i. e.*, to what he terms the capacity of attention. In brief, the method was to compute a fraction, of which the denominator represented the average distance at which the test was visible, and the numerator the mean variation of the several trials. As measured by this arbitrary index, the 10 brightest and the 10 dullest of 300 university students were related, in terms of a common denominator, as 19/1000 and 62.5/1000. Binet (4), however, points out that the dull students had the better eyesight, *i. e.*, the larger denominator, and suggests that the index be taken directly from the mean variation. Abelson, who tested acuity of vision of backward children by a method which afforded five measurements of the distance from the chart moving inward and five moving outward (in each case to the point where the test-object could just be correctly reported), found little or no correlation between imputed intelligence and the average distance (acuity proper), but "the average deviation gave a correlation with the intelligence, the more capable children showing a tendency to small average deviations." He thus confirms van Biervliet.

(7) *Right-eyedness.* While both eyes are employed for binocular vision, there is some evidence that most persons 'favor' one eye, whenever, for any reason, binocular vision is not in use, *e. g.*, in looking through a microscope or telescope. Van Biervliet has measured the visual acuity of 100 persons, whose optical defects had previously been corrected, with the result that the favored eye very uniformly excels the unfavored one in visual acuity by one-ninth: he further asserts that right-handed people are right-eyed and left-handed people left-eyed, and that the same sort of sensorial asymmetry can be demonstrated in audition, cutaneous discrimination, and discrimination of lifted weights.

NOTES.—Statistics of visual defect are rendered difficult of comparison, not only by differences in the methods followed in the investigations, but also at times by failure to state whether examinations were made with or without the glasses actually worn by pupils, or, in case such statement is made, to indicate the effect upon the results of including or excluding trials made

with these glasses. Cohn distinguishes between visual capacity proper (*Sehleistung*) and visual acuity (*Sehschärfe*), which is the efficiency when proper glasses are used. But since numbers of children are daily wearing improperly fitted glasses, one almost needs another term to indicate the vision that is had with these glasses.

REFERENCES

(1) A. R. Abelson, The measurement of mental ability of 'backward' children. *BrJPs*, 4: 1911, 268-314, especially 276 f.

(2) Amy Barrington and K. Pearson, A first study of the inheritance of vision and of the relative influence of heredity and environment on sight. *Eugenics Lab. Memoirs*, 5: 1909.

(3) W. F. Barry, The hygiene of the schoolroom. N. Y., 2d ed., 1904. Pp. 191.

(4) A. Binet, A propos de la mesure de l'intelligence. *AnPs*, 11: 1904 (1905), 69-82.

(5) A. M. Calhoun, Effects of student life upon eyesight. Bureau Ed., Washington, 1881. Pp. 29.

(6) R. B. Carter, Report on vision of children attending elementary schools in London. London, 1896. Pp. 16.

(7) H. Cohn, (*a*) The hygiene of the eye. Eng. tr., London, 1886. (*b*) Die Sehleistung von 50,000 Breslauer Schulkindern, nebst Anleitung zur ähnlicher Untersuchungen für Aerzte u. Lehrer. Breslau, 1899. Pp. 148. (*c*) Was haben die Augenärzte für die Schulhygiene geleistet und was müssen sie noch leisten? Berlin, 1904. Pp. 35. (Contains bibliography of 74 titles.)

(8) G. M. Gould, The cause, nature and consequences of eyestrain. *PopSciM*, 67: 1905, 738-747.

(9) G. M. Gould, Biographic clinics. 5 vols., Phila., 1903-7.

(10) H. v. Helmholtz, Physiologische Optik. 3d. ed.

(11) E. Hope and E. Browne, A manual of school hygiene. Cambridge Eng., 1904. Pp. 207.

(12) L. Kotelmann, School hygiene. Eng. tr., Syracuse, 1889. Pp. 382.

(13) A. Newsholme, School hygiene. Boston, 1894. Pp. 140.

(14) Rep. com. on statistics of defective sight and hearing of public school children. *RepComEd*, 1902, ii., 2143-2155.

(15) S. D. Risley, Weak eyes in the public schools of Philadelphia: rep. com. on examination of the eyes of the children in the public schools of Phila. *Phila.Med.Times*, 11: 1880-1, 673-685.

(16) S. D. Risley, School hygiene, being pp. 353-418 in vol. 2 of Wm Norris and C. Oliver, System of diseases of the eye. Phila., 1897.

(17) Schmidt-Rimpler, Die Schulkurzsichtigkeit u. ihre Bekämpfung. Leipzig, 1890.

(18) F. W. Smedley, Rept. dept. child-study and pedag. investigation. 46th *An.Rep.Brd.Ed.*, Chicago, 1899-1900. Also *RepComEd*, 1902, i., 1095-1115.

(19) S. Snell, Eyesight and school life. Bristol, Eng., 1895. Pp. 70.

(20) Stilling, Die Kurzsichtigkeit, ihre Entstehung und Bedeutung. *SmAbPdPs*, 7: No. 3, Berlin, 1903.

(21) J. van Biervliet, L'asymétrie sensorielle. *BuAcRoySci*, 34: Série 3: 1897, 326-366.

(22) J. van Biervliet, La mesure de l'intelligence, *JPsPa*, 1: 1904, 225-235

(23) G. T. Welch, Report on the examination of the eyes of the public school children of Passaic, N. J. April, 1896. Also summarized in Ref. 14.

(24) A. G. Young, 7th An. Rep. State Brd. Health, Maine. Augusta, 1892. Pp. 399. See especially 128-131.

TEST 15

Balance and control of eye-muscles: Heterophoria.—Strictly speaking, the examination of the condition of the eye-muscles is a physiological test, but because this condition affects clearness of vision, it may be included here with other visual tests.

Each eye-ball is supplied with six muscles. By their action in varying combination, the eye is moved freely in its bed, somewhat after the fashion of a ball-and-socket joint. Under normal conditions, the balance and the innervation of these muscles are such that both eyes move in concert, *i. e.*, the eye-movements are automatically coördinated for purposes of single vision and the lines of regard are restricted to movements where a common fixation point is possible. In some individuals, however, there exists more or less 'imbalance,' or asymmetry of eye-movement, so that the two eyes fail to 'track,' as it were.

If we consider only the relations of the visual line to one another and neglect paralytic affections of the muscles, we may distinguish between latent tendencies toward asymmetry, or *heterophoria,* and actual or manifest asymmetry, *heterotropia* or strabismus. Following the terminology of Stevens (1), we may define the possibilities as follows: *orthophoria* is a tending of the visual lines in parallelism when the determination is made for a point not less than 6 m. distant; heterophoria is a tending of these lines in some other way under the same conditions. Heterophoria may appear (*a*) as *esophoria,* a tending of the lines inward or toward one another, (*b*) as *exophoria,* a tending of the lines outward or away from one another, (*c*) as *hyperphoria,* a tending of the right or of the left visual line in a direction above its fellow,[1] or (*d*) as tendencies in oblique directions, viz: hyperesophoria and hyperexophoria.

[1]The term does not imply that the line which is too high is at fault, but merely that it is higher. Hence, of course, the lack of necessity for any term to indicate that one line is lower than the other.

The tendencies just described are tendencies only, and are latent or concealed in the ordinary use of the eyes on account of the strong 'desire' for binocular vision. For their discovery, accordingly, it is necessary to resort to means for eliminating, so far as possible, this reflex or automatic correction of the latent tendency. The means most commonly employed, as illustrated in the tests that follow, is the establishment of disparate images on the two retinas.[1]

When binocular vision is not habitually attained, the tendencies above described are no longer latent, but manifest, and heterotropia (strabismus or squint) is the result. Heterotropia may appear as esotropia, converging strabismus, or deviation of the visual lines inward; as exotropia, diverging strabismus, or deviation outward; as hypertropia, strabismus sursumvergens or deorsumvergens; or as compound deviations, termed by Stevens hyperesotropia and hyperexotropia.

The most obvious immediate result of heterotropia is diplopia or double vision, a very annoying, but not usually a permanent symptom, because the person thus affected soon comes to neglect the bothersome image from the 'squinting' eye, and to take account only of that from the 'fixing' eye. In time, there results, usually, a limitation of the movements and of the retinal sensitivity of the squinting eye (exanopsic amblyopia), which is one of the most interesting instances of the loss of function through disuse.[2]

Strabismus and heterophoria are functionally associated with ametropia; in particular, divergent displacement is more apt to be associated with myopia, and convergent displacement with hyperopia, probably as a consequence of the straining after clear vision under the hyperopic handicap.

The chief instruments for the detection of muscular asymmetries are prisms of varying construction, the Maddox rod, and the stenopaic lens. Stevens' phorometer is a device for holding and rotating prisms with accuracy and under optimal conditions. The phoro-optometer is a combination of the phorometer with other instruments, such as the Maddox rod, Risley's prism, etc.

---

[1] This assumption that voluntary attempts at fusion will be renounced if the two images are sufficiently disparate, is not entirely correct, and in so far, it is not always possible to make an accurate determination of heterophoria, particularly when slight, by means of the principle of diplopia. Slight heterophoria, moreover, is not to be regarded as abnormal.

[2] As this is particularly to be feared in the case of children, whose eyes have not reached functional maturity, prompt medical attention to strabismus is highly imperative.

Two tests are here detailed, the Maddox rod and the prism test. Both are convenient, portable, and inexpensive, but possess the disadvantage common to all tests for heterophoria held close to the eye, viz: that *S* does not always completely renounce the fusion-impulse.

In the Maddox test, the so-called 'rod' transforms for one eye the flame of a candle into a long narrow streak of red light, while the other eye sees the candle flame naturally. Heterophoria is indicated by the lack of coincidence in these two images.

The prism test, which is essentially an auxiliary test, consists in producing artificial displacement of images by means of the prisms, and measuring *S*'s ability to produce voluntary fusion of these displaced images.

A. THE MADDOX ROD TEST

Apparatus.—Maddox multiple red rod (Fig. 37). Trial frame (Fig. 35). Candle. Meter stick. [A set of trial prisms may be added.]

Preliminaries.—Place the lighted candle on a level with *S*'s eyes and 6 m. distance, preferably in a darkened room. Adjust the trial frame.

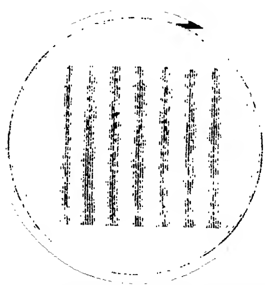


FIG. 37. MADDOX MULTIPLE ROD.

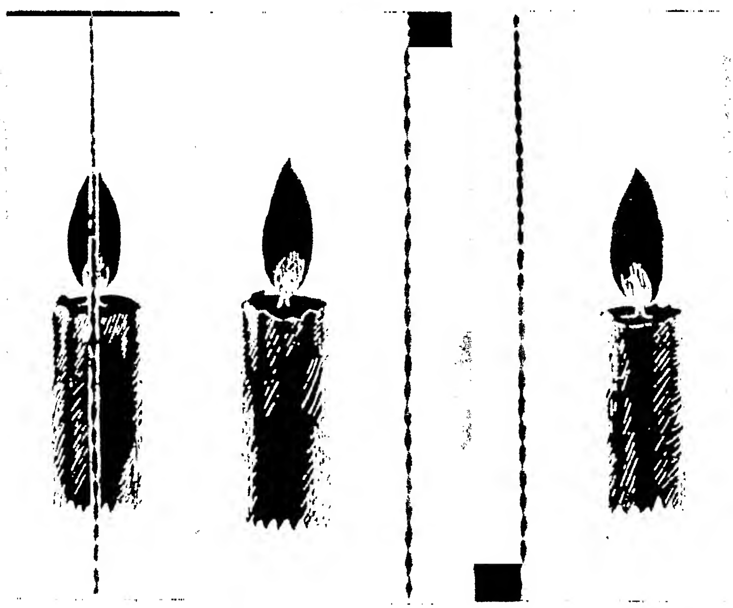
Method.—(1) Let *S* close his left eye: place the Maddox rod in the frame before the right eye with the bars set horizontally. *S* should then perceive a long, narrow, vertical streak of red light. Then let *S* open his left eye and at once state whether the red streak passes exactly through the candle flame.

(2) Turn the rod until the bars run vertically. *S* will see a horizontal red streak. Let him open his left eye and at once state whether the streak passes exactly through the candle flame.

Results.—In the first test, the possible results are: (*a*) the line passes through the flame, orthophoria (Fig. 38); (*b*) the line passes to the right of the flame, esophoria or homonymous displacement (Fig. 39); (*c*) the line passes to the left of the flame, exophoria or crossed displacement (Fig. 40).

In the second test, the possible results are: (*a*) the line passes through the flame, orthophoria (Fig. 41); (*b*) the line passes below the flame, right hyperphoria (Fig. 42); (*c*) the line passes above the flame, left hyperphoria (Fig. 43).

Notes.—Next to orthophoria, esophoria is the most common condition. Unequal vertical adjustment, hyperphoria, is not



FIGS. 38-40. ILLUSTRATING ORTHOPHORIA, ESOPHORIA, AND EXOPHORIA, RESPECTIVELY.

As revealed by the Maddox rod when used before the right eye for horizontal deviation. (De Schweinitz and Randall.)

common, save that an upward deviation of the squinting eye is almost always associated with high degrees of convergent strabismus.

If the latent asymmetry is but slight, there may appear a more or less rapid corrective movement: *S* will then notice lack of coincidence of the line and the flame when the left eye is opened, but the two images soon fuse together. On the other hand, if the asymmetry is larger, *E* may determine its degree by placing prisms before the left eye and ascertaining by trial how strong a prism is needed to enable fusion to occur.

If both horizontal and vertical imbalance is observed, the defect is hyperesophoria or hyperexophoria. This may be demonstrated, if desired, by placing the Maddox rod in an oblique position.

FIG. 41. MADDOX TEST FOR VERTICAL DEVIATION; ORTHOPHORIA.

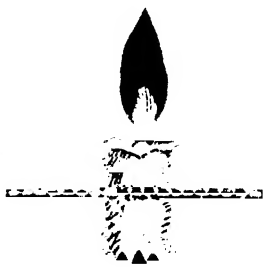


FIG. 42. MADDOX TEST FOR VERTICAL DEVIATION; RIGHT HYPERPHORIA.

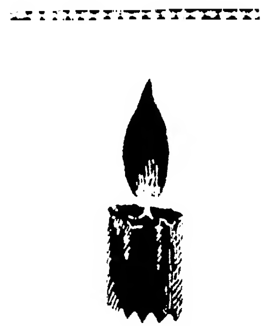


FIG. 43. MADDOX TEST FOR VERTICAL DEVIATION; LEFT HYPERPHORIA.

Stevens' stenopaic lens (Fig. 44) may be substituted for the Maddox rod. A single determination then suffices for both horizontal and vertical displacement. In orthophoria, the candle flame appears in the center of a diffused disc of light; in heterophoria, it is displaced to the right or left, above or below, or obliquely, in a manner corresponding to that of the Maddox line-and-flame test (Fig. 45). The stenopaic lens con-

sists of a convex lens of 13 D., covered, save for a very small opening in the center. The principle is again that of disparate images.

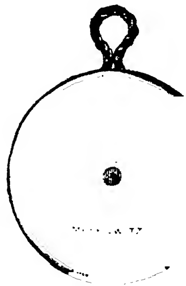


FIG. 44. STEVENS' STENOPAIC LENS.

B. TEST WITH PRISMS

APPARATUS.—Trial frame. Four prisms, one of 2, one of 8, and two of 20 prism-diopters, of the circular pattern fitted for the trial frame.[1] Candle. Meter stick.

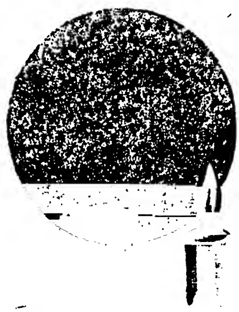


FIG. 45. HETEROPHORIA, AS REVEALED BY THE STEVENS LENS.

[1]These prisms permit *E* to test *S*'s ability to overcome the degrees of displacement that are considered standard for the three positions: their cost is about $7. For a little more money, however, a fairly complete set of prisms may be purchased, which will permit a more flexible test. The strength of prisms is now commonly indicated in prism-diopters. One diopter is a strength of prism which will deflect a beam of light 1 cm. on a tangent plane placed at a distance of 1 m. In the smaller powers this unit is practically identical with degree of prism, or angular size of prism—the unit formerly used.

PRELIMINARIES.—Place the lighted candle on a level with *S*'s eyes and 6 m. distant, preferably in a darkened room. Adjust the trial frame.

METHOD.—(1) To test abduction, or *S*'s ability to overcome a standard amount of displacement by rotating the eyes outward, place the 8-D. prism before one eye with the base in, *i. e.*, toward the nose.

(2) To test *S*'s ability in adduction, or forcible convergence, place a 20-D. prism, with the base-out, *before each eye.*[1]

(3) To test *S*'s ability in sursumduction (compensation for vertical displacement), place the 2-D. prism, with the base either up or down, before one eye.

RESULTS.—With orthophoria, *S* should secure fusion under the conditions imposed, if not at the first trial, at least after a few trials on different days. Failure to accomplish this, or ability to overcome larger angular displacements than those cited, is indicative of heterophoria, or of other inequalities in the set of the eye-balls, *e. g.*, declination.[2]

NOTES.—This test may, of course, be applied to cases in which either orthophoria or heterophoria is present. It may be of value in measuring *S*'s control of his eye-muscles, not only as a matter of optical hygiene, but also in conjunction with tests and experiments of a psychological nature, *e. g.*, stereoscopy, binocular fusion, and visual space-perception in general.

REFERENCES

(1) G. T. Stevens, A treatise on the motor apparatus of the eyes. Phila., 1906. Pp. 496.

(2) W. N. Suter, The refraction and motility of the eye. Phila., 1903. Pp. 390.

TEST 16

Color-blindness.—This test continues the examination of the functional efficiency of the eye as a sense-organ. It has obvious practical import, as well as high theoretical significance in connection with the theory of vision.

---

[1]The ability to overcome prisms by convergence is about 50 D., according to Stevens, but an exact standard cannot be stated.

[2]For further details of this, and other forms of prism test, consult Suter and Stevens.

It is probably safe to say that no theory of vision has thus far been proposed that satisfactorily explains all the known facts. Certainly, in the case of color-blindness, the work of the last few years has disclosed forms and varieties of defective color-vision that offer the greatest difficulty in classification or explanation. While, then, the tests outlined below are satisfactory for most practical purposes, they need for thorough scientific examination to be supplemented by various other methods,[1] and the account of color-blindness here presented is confessedly provisional and abbreviated.

We know that the retina of the normal eye is not equally sensitive to color stimuli in all portions. If small bits of colored paper are moved inward and outward between the periphery of the field of vision and the area of clearest vision, there can be made out three fairly clearly defined zones of different sensitivity—an inner efficient zone, over which we see all colors; a middle zone, over which reds and greens are seen with difficulty or with much altered tone, and an outer zone, over which we commonly see nothing but blacks, whites and grays. If, however, the conditions are altered, especially by exposing colored stimuli of high intensity for a brief period, the boundaries of these zones are decidedly altered: green does, indeed, suffer in peripheral vision, but red, under the proper conditions, may be seen, at least as reddish, as far outward as an object may be seen at all.

It has been customary, and it is tempting, to regard color-blindness as a congenital arrest of development (or in some cases as an acquired, pathological modification) of these zones. If we accept the indication of the experiment with simple colored papers, we should expect to encounter total color-blindness (lack of both central and of intermediate zone) or partial color-blindness of a red-green type (lack of central zone). We should not expect to find blue-yellow blindness, at least as an arrest of development, if it be assumed that the red-green zone represents the latest and hence most unstable development in the structure of the retina.

[1] See, for example, the several methods used by Hayes (8) and by Guttmann (7) in their examination of a number of interesting cases.

As a matter of fact, *total color-blindness,* while well authenticated, is rare, and is presumably a pathological defect. It is accompanied by a distinct reduction in visual acuity, by nystagmus, photophobia, and other disorders of the visual organ.

Again, *blue-yellow blindness* (also termed violet-blindness and *tritanopia*), is in much dispute. It certainly is rare, and probably often pathological in character, though cases have been reported of late which seem to be congenital in origin.[1]

Taking, then, this interpretation of color-blindness, which accords with the hypothesis, typified in the Hering theory of vision, that the components of the retina which respond to the stimulation of light waves are arranged in three pairs—black-white, blue-yellow, and red-green—we would be left with red-green blindness as the typical and characteristic form of partial color-blindness. Deficiency in red would be accompanied by a corresponding deficiency in the complementary green, and conversely. In actual vision, certain reds and certain greens would appear neutral or gray, while stimuli that to the normal eye would be described as red or as green conjoined with blue or yellow would be seen as bluish or yellowish, and the spectrum would thus be divided into a long-waved yellow and a short-waved blue section. That many of the color-blind do conform to these theoretical expectations is not to be doubted, but recent investigations have at least raised the question whether such cases are not to be regarded as limiting and extreme, rather than as typical forms of deficiency (Hayes, 8).

The current terminology of color-blindness is, however, based upon the Helmholtz theory of color vision, wherein the three primary visual and retinal elements are assumed to be red, green and violet. In theory, on this basis, it is evident that an eye might possess all three, or but two, or but one of these visual elements; that, in other words, an eye might possess *trichromatic, dichromatic* or *monochromatic* (*achromatic*) vision. The terms red-blindness, green-blindness and violet-blindness, which would naturally be implied by this theory (on the supposition that some of the three primary components

[1]For examples, see Richardson.

were wanting) have, however, been discarded by all careful writers because they are misleading. In their place are employed the terms *protanopia, deuteranopia* and *tritanopia,* meaning the existence of a defect in vision allied to disturbance in the function of the first, second or third of the three components, respectively.

To comprehend the inter-correspondence of nomenclature, it must be understood that the ordinary red-green blindness of the Hering terminology appears as a rule in one of two sub-types, though the distinction is of more theoretical than practical significance. Those belonging to the first type locate the brightest part of the spectrum, as do normal persons, in the yellow; they are called deuteranopes (or rather erroneously green-blinds). Those who locate the brightest part of the spectrum in the yellow-green region and see the entire blue end of the spectrum relatively brighter than the normal person are called protanopes (or erroneously red-blinds). Protanopia is far less frequent than deuteranopia.

Of great theoretical interest and decided practical importance are those cases in which the same color qualities exist as in normal eyes, but in an unusual or anomalous form, whose vision is, therefore, trichromatic, but abnormal. These cases are known popularly *color-weaks* and technically as *anomalous trichromates.* According to Guttmann, anomalous trichromasy manifests seven characteristic symptoms—a reduced sensitivity to color stimuli, especially when the stimuli are of short duration, small area, or low intensity, decidedly heightened color contrast, difficulty in comparing color tones of unequal brightness or saturation and quick fatigue to color stimuli. While all these symptoms, taken together, constitute a characteristic symptom-complex, the reciprocal interaction of the several symptoms varies greatly in individuals, so that the defect actually shows itself in protean forms, and there exist all degrees of anomalous trichromasy, ranging from almost normal vision to practically dichromatic vision.[1]

Here, again, we meet terminological difficulties. The terms color-weak, green-weak and red-weak, which are advocated by Guttmann and employed by some investigators, are opposed by others, *e. g.*, by Nagel, who prefers the term anomalous trichromasy to cover the group as a whole and protanomaly and deuteranomaly to distinguish the two main sub-types, analogous to the two main sub-types of dichromasy. The term 'anomaly,' he argues, indicates a condition of color sensitivity variant from the normal, without implying that the deficiency need be a weakness. This term is better than more specific terms like 'red-weakness' or 'green-weakness,' for the 'weakness' of both of these forms actually shows most in the greens and violets and far less strikingly in the reds.

[1]On color-weakness, consult especially Rosmanit, Nagel and Guttmann.

Holmgren contrasted total color-blindness with partial color-blindness, and divided the latter into complete partial color-blindness and incomplete partial color-blindness (confusions with the green, but not with the red test skein). This division has not been often used, but the term 'color-weakness' has been extensively employed in place of Holmgren's incomplete partial color-blindness, though not quite correctly, because this group, as determined by the Holmgren test, may embrace both von Kries's deuteranopes and the so-called color-weak.[1]

The use of confusing terminology, however, is far less serious than other errors which are exhibited in texts descriptive of color-blindness. The reader may consult, for instances, a book by Abney (1), which embodies his Tyndall lectures of 1894, and a magazine article by Ayers (2). In Abney there will be found a colored frontispiece, taken from the Report of the Royal Society Committee on Color Vision in 1892, which purports to show the spectrum as seen by the color-blind. The spectrum is shown in green and blue: what becomes of yellow is not explained. In Ayers' article there will be found some very pretty colored pictures of roses and Venetian scenes as observed by the color-blind,—pictures that are good examples of the illustrator's art, but absolutely false examples of color vision. Mrs. C. L. Franklin (6) has charitably applied the term "pseudo-scientific" to such writing. A more nearly correct representation of the spectrum seen by the color-blind is given by Thomson (20).

One of the best illustrations of anomalous trichromasy is afforded by the Rayleigh test-equation. Lord Rayleigh (17) discovered that when a homogeneous yellow is equated with a mixture of red and green, the proportions of red and green needed to balance the yellow show only slight variations for the great majority of persons, but that there exists a distinct type of eye that demands a distinctly different combination: most often the mixture must be made much greener, less often much redder. The first type indicates green-anomaly (green-weakness, deuteranomalous trichromasy), the second red-anomaly (red-weakness, protanomalous trichromasy).

While, in theory, the 'color-weak' are not to be identified with the color-blind, for practical purposes it is equally important to diagnose the existence of the defect. If, as seems probable, anomalous trichromasy is as prevalent as true dichromasy, then, since many of the commonly used tests for color-blindness fail to detect color-weakness, and since conditions of daily life would often demand color discrimination under unfavorable conditions, *e. g.,* in railway and marine service,

[1]An interesting tabular summary, confessedly provisional, of the entire field of color defect has been worked out within the compass of a single page by Nagel (16c, p. 308): this the reader is advised to consult before attempting any extensive reading on the theory of color-blind-

it would seem very probable that some disasters may be traceable to this deficiency, which has escaped detection by medical examiners.

Thus, in Germany, among 1778 members of railway regiments, all of whom had passed the wool test and many of whom had also passed Stilling's test, 13 dichromates and 31 anomalous trichromates of various types were discovered by the use of Nagel's test in the hands of military physicians.[1]

Color-blindness may be binocular or monocular. The latter is rare, but naturally of great theoretical importance in determining the nature of color-blindness.

Color-blindness is usually congenital, and then incurable. The common form, red-green blindness (dichromasy), is regarded as an arrest of development, usually, or reversion to a more primitive form of retina. All acquired cases, variously attributed to traumatism, neuritis, atrophy of the optic nerve, hysteria, excessive fatigue, over-indulgence in tobacco, are accompanied by lessened visual acuity, are pathological, and of relatively small concern to the theory of color-vision.[2] Acquired red-green blindness, according to Köllner, is a middle stage between normal vision and total color-blindness, and indicates some grave disorder. The disturbance of function may be located either in the retina or in some part of the nervous system concerned in vision.

Color-blindness seems to have been first noted in literature in 1684, but first described accurately by Dalton, the celebrated English chemist, in 1794. The first attempt at a systematic examination of a large number of cases was made by Seebeck in Berlin in 1837 by the aid of colored papers. The first systematic examination of railway employees dates from 1875, when a serious accident in Sweden led Holmgren (11), of the University of Upsala, to devise his well-known wool test and to induce officials to adopt it.[3]

The chief devices and methods for testing color-blindness are Holmgren's, Galton's, Thomson's, Oliver's, and other assortments of colored worsteds, Stilling's pseudo-isochromatic charts, Nagel's card test, spectroscopic examination, various contrast tests, and the use of equations of mixed colors, particularly Nagel's equation-apparatus, and Hering's apparatus, which enables the examiner to adjust a color equation of transmitted light that shall appear to the color-blind as uniform gray. In addition, numerous forms of color-blindness lantern (Williams', Friedenberg's, Oliver's, etc.) have been devised for testing railroad and marine employees by simulating the con-

[1]For summaries of recent discussions of the practical dangers of color defects, proposals for improving tests, for altering signal lights, etc., consult Hayes (9). An annual review of the literature of color defects may be expected in the *PsBu*, March issues.

[2]For illustrative cases, consult Collin and Nagel, and Köllner.

[3]For other details of the history of color-blindness as well as a discussion of methods, though not brought down to date, consult Jennings (12) and Thomson (20).

ditions of night-signalling, and soiled signal-flags have been used for similar purposes, while Henmon has proposed a discrimination-time test.

Two forms of test are here described: the familiar and widely used Holmgren wool test, adopted by the American Ophthalmological Society, and Nagel's new card test, which is specially fitted for the diagnosis of color-weakness and of other variant types of defect. Both of these tests are inexpensive, compact, and portable. They may be employed in conjunction with one another, but the Nagel test is undoubtedly the better of the two.

It is well to point out that, by dint of daily experience, the color-blind individual develops a capacity to recognize some reds and greens by means of secondary criteria, such as brightness (tint) and saturation (chroma), and familiarity with the application of color nomenclature (grass is green, cranberries red, etc.), so that the defect may exist unrecognized, either by himself or by his acquaintances, until chance compels the recognition or discrimination of tones to which these criteria cannot be applied. Hence arises the necessity, in the administration of tests, of displaying a large number of colors of varied saturation and brightness, in order that, for any individual, some combination or series of combinations of colors may be found, in the recognition of which these criteria cannot be used. Here, too, appears in some part the explanation of the seeming individuality of defective color-vision.

A. THE HOLMGREN WOOL TEST

MATERIAL.—Holmgren's worsteds.[1] Sheet of light gray or white cardboard or a similarly colored cloth.

METHOD.—(*a*) *Full procedure.* (1) Remove the three large test skeins, pale green, rose and red, Nos. A, B, C. Scatter the remaining skeins over the cloth or paper in *diffuse daylight* only.[2] Hand to *S* the green test skein, No. A, and direct him to pick from the table all those skeins that resemble the test skein, *i. e.*, all the tints and shades of that color. Explain that there are no two specimens alike, and that an exact match is not required. It will do no harm to illustrate the process by

[1]Examination of several boxes of test-worsteds supplied by different dealers in optical supplies shows that in many of them the skeins diverge widely in color-tone and saturation from the original Holmgren skeins. When this variation from standard affects the standard test-skeins, the whole set of worsteds evidently becomes valueless. The C. H. Stoelting Co. is now in a position to supply skeins in the exact original tones selected by Holmgren.

[2]When not in use the skeins must be carefully enclosed in their box, as they will fade or change color if continuously exposed to light.

selecting two or three skeins for him, provided these are afterward mixed with the pile. To save time in explanation, other *S*'s may be allowed to watch this demonstration.

(2) If hesitation appears, or if grays, browns or reds as well as greens are selected, continue the test by the use of the rose skein, No. B. The typical color-blind will then select some blues or purples, or, less often, grays or greens.

(3) Finally, the red test skein, No. C, may be used, though many color-blinds have little difficulty with this test on account of the strong saturation of the test skein.

In all three tests, preserve a careful record of the skeins selected by *S*'s who deviate in any particular from the normal.

(*b*) *Abbreviated procedure.* This test may be used for quick preliminary examination. Place irregularly on the cloth four green skeins (*e. g.*, 1, 11, 31, 65) and eight 'confusion' skeins of gray, brown, and pink (*e. g.*, 20, 98, 118, 59, 189, 199, 13, 23). Hand to *S* the pale green standard, No. A, and require him to pick from the cloth as rapidly as possible four skeins that match the test skein (in the sense previously described). Allow him approximately 4 sec. to make this selection. If this test can not be promptly and accurately executed, examine *S* further by the full procedure.

TYPICAL RESULTS.—(1) About 4 per cent. of men and less than 0.5 per cent. of women are color-blind: the most common defect is probably red-green blindness of the form known as deuteranopia.

(2) The following are actual selections of a typical red-green blind. By assembling these skeins *E* can gain the best idea of the nature of the confusions likely to be discovered.

Green standard: 1, 11, 51, 61, 10, 20, 17, 98, 118, 59, 119, 179, 189. (Occasionally some pink, like 12, is also selected.)

Rose standard: 32, 43, 113, 123, 104, 114, 124, 6, 46, 56, 67, 77, 87, 18.

Red standard: 33, 43, 53, 63, 73, 83, 133, 54, 49.

NOTES.—Inability to name colors rightly has sometimes been erroneously mistaken for color-blindness, but the term must be applied only to instances of actual inability to *see* colors

rightly. Consequently, no color-blindness test should *hinge* upon the ability to name correctly the various colors presented, and, in the conduct of the Holmgren test at least, reference to color names should be avoided if possible.

If *S* works very slowly and hesitatingly, but finally makes a correct selection, this may indicate several possibilities, which should be tested by further study of the case. (*a*) The slowness may be due merely to extreme cautiousness on his part, coupled with some anxiety or uneasiness about the test, or with failure to understand clearly just what is wanted. (*b*) The slowness may be due, in the case of very young children or untutored adults, to gross ignorance of, and unfamiliarity with colors. (*c*) The slowness may be indicative of color-weakness, in which case Nagel's test should be applied for further diagnosis of the defect.

All instances in which specific color differences are at first recognized with difficulty or not at all, but in which, after coaching or instruction, an efficiency is developed adequate for passing the test in use, must be looked upon with suspicion, and it must not be assumed forthwith that color-blindness has been cured by training, for either the cautiousness or ignorance just mentioned were present at first and removed by the training, or the conditions of the test were too simple, and secondary criteria were developed by *S*.

If apparent cases of blue-yellow, or of total color-blindness are discovered, these should, if possible, be given most careful examination by an expert psychologist.

It is obvious that many callings are, or should be, closed to the color-blind, *e. g.*, railroading, marine and naval service, medicine, chemical analysis, painting and decorating, certain branches of botany, microscopy, mineralogy, the handling of dry goods, millinery, etc. In some phases of school work, the color-blind pupil is likewise at an evident disadvantage.[1] The test should, accordingly, be regularly instituted in the early years of school life, in order that the existence of the defect may be made known to the child as soon as possible.

[1]On the consequences of color-blindness for daily life, see especially Jerchel (18).

B. THE NAGEL CARD TEST

MATERIAL.—Nagel's color-blindness cards, 7th edition.[1]

METHOD.—A. Spread out the 16 cards of Section A upon a table. It is imperative that there be good daylight illumination and that *S* stand upright before the table and not bend over it or pick up the cards to examine them at close range. Provide *S* with a pen-holder or other convenient pointer by which he may indicate his responses. Ask him the following questions:

1. Can you point out to me some cards on which *red or reddish* spots are to be seen?
2. On which cards are there *red spots only?*[2]
3. On which *green spots only?*
4. On which *gray spots only?*

B. Remove the cards and show *S*, one at a time, the four cards of Section B: ask him for each card what colors he sees.

Responses of normal S's. The questions of Section A are answered rapidly and correctly, save in the case of a few timid, anxious or unintelligent persons, and these will usually reply correctly to the diagnostically more important questions, Nos. 3 and 4, once they have learned what is wanted, especially that they need not find a whole ring that is red, etc., but only single spots. A wrong answer that is afterward spontaneously corrected is not symptomatic of defective color-vision.

In dealing with Section B, *E* must remember that he is not here testing for fine discrimination in the use of color nomenclature, but merely determining whether or not *S* recognizes red and green as such, and whether he designates as red or as green some color which is neither red nor green. Thus, in Card B 1, a normal *S* readily recognizes the green, but may hesitate to name the yellow-brown, and it is indifferent whether he calls it brown, yellow, or even gray. But it would be very suspicious if he called all the spots green, or if he called the brown spots red, reddish or orange. Similarly, if the brown spots on B 2 and

[1]Those who desire to conduct special supplementary tests for blue-yellow blindness, or for differentiating between protanopia and deuteranopia, etc., will find that the earlier editions of Nagel's test make special provision for this work. Cf. the directions printed in the first edition of this *Manual* to accompany the 4th edition of the Nagel test.

[2]Do not ask for bluish red or rose-colored spots.

B 4 are called green, *S* is surely defective in color vision, and if all the spots are called red, he is certainly a color-blind, whereas, again, if the brown spots are called yellow or even gray, he need not be color-blind.

Responses of S's with defective color-vision. Errors are evident, as a rule, even in the first question of Section A. Most of the color-blind select as containing red spots Cards 6 or 11 or 12, on which they mistake the yellowish-green or the brown spots for red. Unless there be deliberate simulation, these errors are well-nigh complete evidence of color-blindness. The anomalous trichromates usually answer this question without mistake, though they may show hesitation in their responses.

With Questions 2, 3 and 4, the color-blind are reduced to sheer guess-work. The anomalous trichromates can answer Question 2 (red spots only) correctly, but with Questions 3 and 4 they will show decided uncertainty and slowness, and will, indeed, not seldom answer both incorrectly by designating cards that contain both green and gray when asked for cards containing only green and only gray. Occasionally, too, they may select cards that contain some red spots.

In Section B, color-blind *S*'s see but a single color on each card and the color-names they select are a matter of pure guess-work. B 1 is often called green, B 2 and B 4 red, B 3 gray or green. The anomalous trichromates see as a single color only B 1 and, less often, B 3; they see B 2 and B 4 plainly in two colors, although they often call the brown green by contrast with the adjacent red.[1]

Any *S*, therefore, who sees the red and brown cards, B 2 and B 4, as a single color, or as red and green, is certainly defective in color-vision. But an error with B 1 only or with B 3 only merely raises a suspicion of defective color-vision, and does not permit a positive diagnosis of deficiency if *all* the other questions are correctly answered. To justify the suspicion of defect, such an *S* must respond incorrectly or hesitatingly to Questions 3 and 4.

[1]Normal *S*'s may hesitatingly admit that the brown spots look somewhat greenish, but they never deliberately describe these two cards as composed of red and green spots.

NOTES.—The cards should not be placed under glass to prevent contact with them, since this introduces a disturbing reflection. If *E* wishes to put the cards in an album or otherwise arrange them to obviate spreading them out upon a table at each trial, he should take care to present them in different positions, so that it will be impossible for *S*'s to learn their position or to communicate them to other possible examinees. If anxiety or ineptitude on the part of *S* renders the diagnosis doubtful in the first trial, it is best to forego further trial until a later day and then to try the whole test over again.

It is imperative that *E* adopt a quiet, sympathetic manner, free from any sign of irritation or impatience, especially when dealing with slow or stupid *S*'s, or even with those who are plainly attempting deceit. During the test, *E* must avoid informing *S*, whether directly or by suggestion, of any mistakes he has made. Discussion or criticism of *S*'s selections is out of place, while, for the sake of future tests, it would be desirable not to explain *S*'s errors to him in detail, even after the test.

If *S* has decidedly low visual acuity, this must be corrected, at least approximately, by appropriate lenses, before the color-blindness test is begun.

REFERENCES

(1) W. Abney, Color vision. London, 1895. See also his Researches in colour vision and the trichromatic theory. London, 1913. Pp. 423.

(2) E. Ayers, Color-blindness, with special reference to art and artists. *Century Mag.*, 73: 1907, 876-889.

(3) J. W. Baird, The problems of color-blindness. *PsBu*, 5: 1908, 294-300.

(4) Collin and W. Nagel, Erworbene Tritanopie. *ZPs*, II Abt., 41: 1906, 74-88.

(5) Dalton, Extraordinary facts relating to the vision of colors. Trans. of the Lit. and Philos. Soc. of Manchester, 1794.

(6) Mrs. C. L. Franklin, Magazine Science. *Sci*, n.s. 25: 1907, 746.

(7) A. Guttmann, Untersuchungen über Farbenschwäche. *ZPs*, II Abt. 42: 1906, 24-64, 250-270; 43: 1909, 146-162, 199-223, 255-298.

(8) S. P. Hayes, The color sensations of the partially color-blind: a criticism of current teaching. *AmJPs*, 22: 1911, 369-407.

(9) S. P. Hayes, Vision: Color defects. *PsBu*, 10: March, 1913, 101-107. (Similar annual summaries in previous years *PsBu*.)

(10.) V. Henmon, The detection of color-blindness. *JPh*, 3: 1906, 341-4.

(11) Holmgren, Die Farbenblindheit in ihrer Beziehung zum Eisenbahn- und Marine-dienst. (German edition.) Leipzig, 1878.

(12) J. E. Jennings, Color-vision and color-blindness. Phila., 1896. Pp. 111.

(13) W. Jerchel, Inwieweit wird das Medizinstudium durch Rotgrünblindheit beeinflusst? *ZPs*, II Abt., 47: 1912, 1-33.

(14) H. Köllner, (*a*) Die diagnostische Bedeutung der erworbenen Farbensinnstörungen. *BerlinKlW*, 48: 1911, 846-9, 897-900. (*b*) Die Störungen des Farbensinnes, ihre klinische Bedeutung und ihre Diagnose. Berlin, 1912. Pp. 428.

(15) J. v. Kries, (*a*) Ueber Farbensysteme. *ZPs*, 13: 1897, 241-324. (Corrective note, p. 473.) (*b*) Die Gesichtsempfindungen. Nagel's Handbuch der Physiologie des Menschen. Braunschweig, 3: 1905, 109-282. (*c*) Normale und anomale Farbensysteme. Helmholtz's Handbuch der physiologischen Optik, 3te Aufl., 2: 1911, 333-378.

(16) W. Nagel, (*a*) Fortgesetzte Untersuchungen zur Symptomatologie u. Diagnostik der angeborenen Störungen des Farbensinns. *ZPs*, II Abt., 41: 1906, 238-282, 319-337. (*b*) Zur Nomenclatur der Farbensinnstörungen. *Ibid.*, 42: 1907, 65. (*c*) Ueber typische und atypische Farbensinnstörungen. *Ibid.*, 43: 1908, 299-309. (Consult also other articles by this writer in the same periodical.)

(17) Rayleigh, Experiments on colour. *Nature*, 25: 1881, 64-66.

(18) Florence Richardson, An unusual case of color-blindness. *PsBu*, 8: 1911, 55-6. (Also summary of another case, 214-5.)

(19) J. Rosmanit, Zur Farbensinnprüfung im Eisenbahn- und Marinedienste. Vienna and Leipzig, 1907. Pp. 59.

(20) W. Thomson, Detection of color-blindness, being pp. 315-352 in Norris and Oliver, System of diseases of the eye, vol. ii, Phila., 1897.

TEST 17

Discrimination of brightness.—The object of this test is to obtain an index, for comparative purposes, of *S*'s ability to distinguish very small differences in brightness, or more exactly, to determine the smallest difference in brightness that *S* can distinguish under simple experimental conditions. The present test omits consideration of chromatic stimuli, and is confined to the discrimination of brightness, first by the use of reflected, secondly by the use of transmitted light.

Visual discrimination has been studied in the laboratory by many competent investigators, *e. g.*, Ament, Aubert, Bouguer, Helmholtz, Fröbes, Kraepelin, Masson, Merkel, Schirmer, Volkmann, and others. Tests of school children by Gilbert (3) and Spearman (6) have followed simpler methods.

In the laboratory, use has been made of Masson's disc, both by daylight and artificial illumination, of the episkotister, of gray papers and of shadows. Toulouse (9) proposes solutions of aniline colors in glass receptacles. Gilbert used a series of ten pieces of cloth soaked in a red dye of graded intensity. Investigations that are most comparable with the method here proposed are those of Ament (1), Fröbes (2), and Spearman,

all of whom made use of gray papers, and of Gilbert, who examined school children, though with chromatic stimuli.

A. DISCRIMINATION OF GRAYS—REFLECTED LIGHT

APPARATUS.—Set of 10 test-cards, each composed of two gray strips, 13 x 40 mm., on a white background, 10 x 10 cm. Exposure frame, fitted with a card-holder which may be rotated

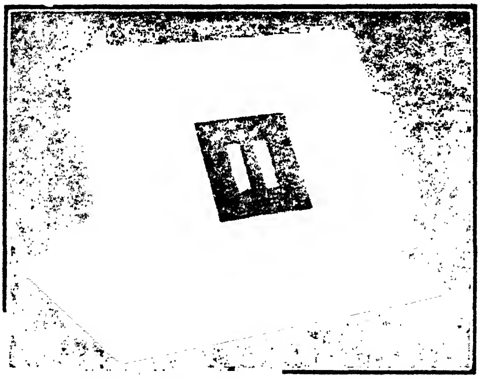


FIG. 46. APPARATUS FOR THE DISCRIMINATION OF GRAYS.

through 180°, and with a black screen, through an opening (8 x 8 cm.) in which the test-cards may be viewed (Fig. 46). Light gray cloth, about 70 x 160 cm., for a background. Two supports, with angle-pieces, and a horizontal rod 70 cm. long. Headrest.

The cards are numbered from 0 to 9, corresponding to 10 different pairs of stimuli. Each card contains one strip of the lightest or standard gray, and one strip of comparison gray. Card No. 0 represents no difference, or objective equality; Card No. 1 represents the minimal objective difference; Card No. 9 the maximal objective difference and is easily supraliminal for the normal eye. Each card is numbered on the back in such a way that, when looking at the face of the card with the number up, the right strip is the darker; there is also a small black mark on the extreme edge of the card on the side of the darker strip.

The grays used on these cards have been specially prepared, under the author's direction, by S. L. Sheldon, photographer, of Ithaca, N. Y., and have been carefully standardized. Each set of grays is printed from the same negative, on which the original series was formed by graded serial exposures before a sheet of milk glass set in a north window. They will not fade or change their tone, unless brought into contact with chemical fumes or solutions; but, for additional protection, they should be kept under cover when not in use, and never be handled in bright sunlight.

The tones, sizes, and spatial relations of the strips, cards, and background have been selected to eliminate errors that might arise from

adaptation and contrast. The size of the strips is slightly smaller than that used by Ament (18x45 mm.) and slightly larger than that prescribed by Titchener (10x40 mm.) for the demonstration of Weber's law in brightnesses.

PRELIMINARIES.—Place a small table, say 65 x 90 cm., squarely before a window where good diffuse daylight may be secured (preferably a north window with full clear exposure to the sky); leave just enough space between the front of the table and the window for two chairs for *S* and *E*. Spread the gray cloth over the top of the table, and stretch it up vertically at the back edge by means of the supports, so as to form a continuous background of gray, with the vertical back at least 65 cm. high and about 65 cm. distant from *S*'s eyes.

Place the exposure frame in the center of the table at the optimal reading distance (about 35 cm., unless *S* has uncorrected myopia or hyperopia), and adjust its height so that the *top* of the frame is on a level with *S*'s eyes. Adjust the headrest so that *S* may sit erect, squarely before the exposure frame and close to the table-edge, with his back, of course, to the window.

Keep the test-cards conveniently near, but out of *S*'s sight. *E* will find it most convenient to sit at *S*'s right.

METHOD.—(1) Spend 5 min. in giving *S* practise and familiarity with the test. For this purpose, begin with the large-numbered cards, and pass in general toward the smaller numbers, but without following any rigorous order. With each card, rotate the turn-table, so that the right strip is now the darker, now the lighter: follow an irregular order, and keep *S* always in ignorance of the actual location of the darker strip, and of the correctness of his judgments. In each trial, *S* must report his judgment *in terms of the right-hand strip,* saying either "darker," "lighter," or "equal." (Any doubtful cases may be classed as equal.)

When not observing a test-card, *S* should rest his eyes by directing them toward the gray background. He turns his eyes to the test-card at *E*'s "now," and should be asked to pass a judgment *within 5 sec.* It is not necessary to record results at this point, but from this practise work, *S* will attain a general familiarity with the test, and *E* will form a fair idea of *S*'s 'critical' region.

(2) Proceed now, more formally and exactly, to determine *S*'s difference limen by selecting a stimulus difference which has appeared in the preliminary series to be just noticeable for him. Give this stimulus-card 10 times, 5 times with the right strip darker, 5 times with the right strip lighter, but in chance order.[1] Inform *S* that he will be shown the same card 10 times, but in different positions, of which he is to be ignorant. He must judge either "lighter" or "darker." *S* must not be informed during the series whether his judgments are right or wrong. If *S* gives 8 right answers in 10, the magnitude of the brightness difference then in use affords the desired index.

(3) Confirm the result by testing *S* 10 times with a slightly larger difference, and 10 times with a slightly smaller difference. Unless the tests are disturbed by the operation of such factors as fatigue, loss of interest, practise, fluctuations of attention, etc., *S* may be expected to give 9 or 10 correct judgments in the former, and fewer than 8 in the latter test.

Variations of Method.—Test the discriminative capacity of each eye separately, as well as in conjunction. Employ the trial frame of Test 14, placing the solid disc before the untested eye. Care must be taken to avoid visual fatigue under these conditions. This variation of method is of interest in connection with recent work on psychophysical asymmetry and the relations between right-handedness, right-eyedness, right-earedness, etc. (See, for example, Van Biervliet.)

If means are at hand to secure effective constant illumination by artificial light, this may be tried for comparison with daylight illumination.

Treatment of Results.—For comparative purposes, *S* may be ranked in terms of the arbitrary units afforded by the card-numbers. For more exact quantitative expression, however, the results should be expressed in terms of the brightness-differences which correspond to the card-numbers. This correspondence must be worked out by *E* for the papers employed. Full directions for a simple and sufficiently accurate

[1] It is convenient to prepare on small slips, beforehand, a number of chance orders, and to follow one of these with each set of 10 trials.

photometric determination of brightness values of gray papers will be found in Titchener (Pt. I., 35 ff.).

B. DISCRIMINATION OF BRIGHTNESSES—TRANSMITTED LIGHT

APPARATUS.—Headrest. Brightness discrimination apparatus (Fig. 47). [This is a box fitted with a high power frosted tungsten lamp, the light of which is reflected from two independently adjustable white screens upon two oblong, trans-

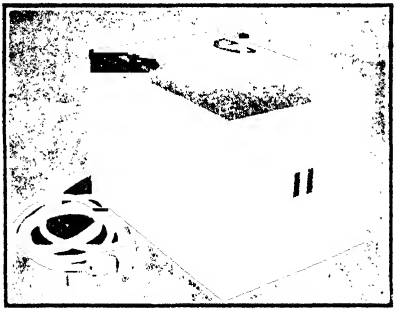


FIG. 47. APPARATUS FOR BRIGHTNESS DISCRIMINATION.

lucent windows, so placed in the face of the box as to give the same dimensions and spatial relations as obtained in the case of the gray strips.]

PRELIMINARIES.—The lamp cord is to be attached to a suitable current (106-110 volts, unless special lamps are ordered). *E* should endeavor to conduct the test in a dark or darkened room. If a brightly lighted room must be used, the effectiveness of the illumination of the 'windows' may be increased by erecting a protecting screen of cardboard or cloth around them.

METHOD.—It is extremely important to arrange the headrest so that *S* is directly in front of the apparatus, with his eyes on a level with the windows in the box. The distance is less important; 50 cm. will be found convenient. The degree of illumination is controlled by two levers, which move the reflecting screens, and which are provided with scales upon the upper surface of the box. *E* first sets the right-hand lever at the point which affords the maximal illumination of the right-hand window, and records the scale-reading exactly. In ac-

cordance with the methods just outlined for the discrimination of grays, *E* now determines the just discriminable difference in the setting of the two levers (when either one of them is at the maximal point). The same precautions must, of course, be taken to reverse the standards in order to correct the space error.[1]

VARIATIONS OF METHOD.—Substitute a 32 C. P. ruby lamp for the frosted lamp, and determine the discriminative capacity for reds of different brightness. Other colors may be employed similarly in this apparatus.

RESULTS.—(1) *Trained observers,* working under conditions similar to those prescribed, can discriminate a brightness difference of 1/120, though this fraction is appreciably altered by changes of technique or of experimental conditions. Untrained observers have less efficiency, about 1/60, according to Spearman. With the reflected light apparatus, most adults can just discriminate Card No. 3.

(2) By a different method and with colored stimuli, Gilbert found that discriminative ability increases very gradually up to the *age* of 17, but exhibits marked irregularities at the age of 7.

(3) In discrimination of shades of color, one may conclude from studies by Nichols (5), Gilbert (3), and Thompson (7) that women and girls very slightly exceed men and boys in this capacity. Luckey, however, concluded that no *sex differences* could be demonstrated in color discrimination.

(4) Individual *S*'s are apt to possess a constant *space error, i. e.,* to tend to judge the gray on one side darker; in some cases this is the right, and in others the left, but it seems impossible to correlate this asymmetry with right and left-handedness (Spearman).

(5) Gilbert found no very decided correlation between *visual discrimination and intelligence.* Spearman's experiments upon 24 village-school children give correlations between brightness discrimination and common sense, school cleverness, and gen-

[1]Since the scales are identical and the entire instrument is symmetrical, a given setting of the lever will produce the same intensity of illumination for either window.

eral intelligence in the neighborhood of $+0.50$. In a series with high-class preparatory-school boys, however, school place and brightness discrimination gave only $+0.13$ for the 'raw' correlation.

NOTES.—It is imperative that the conditions under which the gray strips are observed should be kept as constant as possible. Backgrounds, cards, and holder provide these conditions in part, and relative brightness is not affected within a fairly wide range of illumination: nevertheless, it is desirable to work in the same place, at the same time of day, and under closely similar conditions of outdoor illumination, *e.g.*, between 9 a.m. and 3 p.m. on sunshiny days, and at a north window. To ensure evenness of illumination and absence of any shadows, *E* should test the setting of the experiment by placing Card No. 0 in the holder and reversing its position several times. As this card represents objective equality, any constant judgment of difference may serve to indicate uneven conditions of illumination.

In working with brightness differences, and indeed, with all small differences, *E* must be very careful to avoid suggestion of the direction of the difference to *S*, and must keep a persistent watch for all kinds of secondary criteria of judgment. If desired, one could experimentally determine the degree of objective brightness difference that could be overcome by suggestion.

REFERENCES

(1) W. Ament, Ueber das Verhältniss der ebenmerklichen zu den übermerklichen Unterschieden, etc., *PhSd*, 16: 1900, 135-196.

(2) J. Fröbes, Ein Beitrag über die sogenannten Vergleichungen übermerklicher Empfindungsunterschieden. *ZPs*, 36: 1904, 344.

(3) J. A. Gilbert. (*a*) Researches on the mental and physical development of school children. *SdYalePsLab*, 2: 1894, 40-100. (*b*) Researches upon school children and college students. *UnIowaSdPs*, 1: 1897, 1-39.

(4) G. Luckey, Comparative observations on the indirect color range of children, adults and adults trained in color. *AmJPs*, 6: 1895, 489-504.

(5) L. Nichols, On the sensitiveness of the eye to colors of a low degree of saturation. *AmJSci*, 30: 1885, 37.

(6) C. Spearman. General intelligence objectively determined and measured. *AmJPs*, 15: 1904, 201-293.

(7) Helen B. Thompson, The mental traits of sex. Chicago, 1903. Pp. 188.

(8) E. B. Titchener, Experimental psychology. Vol. II, N. Y., 1905.

(9) E. Toulouse, N. Vaschide, and H. Piéron, Technique de psych. expérimentale. Paris, 1904. Pp. 330.
(10) J. van Biervliet, L'asymétrie sensorielle. *BuAcRoySci*, 34: Série iii., 1897, 326-366.

TEST 18

Auditory acuity.—This test, like that of visual acuity, is primarily conducted for hygienic and practical purposes, especially in the examination of the physical condition of school children, and constitutes the chief auditory test. We may distinguish between simple acuity tests, which are designed merely to detect the existence of lessened aural efficiency and roughly to measure its degree, and more elaborate tests of a diagnostic character, which are for the most part not used in group investigations, but are confined to the work of specialists in otology or in the psychology of audition.[1] Among the latter tests may be mentioned that of binaural pitch-difference, integrity of the tonal scale, bone vs. air-conduction, determination of relative and absolute deafness, diagnostic speech-tests, etc. These tests are designed to investigate the functional efficiency of the various auditory structures, such as the tympanum, ossicles, cochlea, auditory nerve, and to determine the cause of the defect in hearing and the possibility of alleviating it by medical treatment. In particular, it is important, from this point of view, to differentiate between defect in the middle, and defect in the internal ear, because in the former case partial deafness may often be relieved, whereas in the latter medical treatment is ordinarily of no avail.

The more common and widely employed tests for acuity fall naturally into two main groups, viz: speech tests and instrumental tests. Speech tests may be conducted by either vocalized or whispered speech, and by either the method of varied range or the method of constant range. For instrumental tests use is most often made of the watch, of some form of audiometer, or acoumeter, or of a tuning fork. The relative merits of these methods and instruments deserve brief consideration.[2]

[1] A typical illustration is given by the interesting article of Bingham (4.)
[2] For further discussion, particularly of the difficulty of measuring the physical energy of minimal tones and noises, see Pillsbury (11).

A. Methods. In the use of both speech and instrumental tests it has been customary to employ the *method of varied range.* A range line is chalked off on the floor of the room; *S* is seated at one end of this range, while *E* moves methodically forward and backward over it, until he determines the extreme limit of auditory capacity for the voice or instrument. The assumption that is made here, viz., that the intensity of a sound decreases as the distance from its source increases (approximately as the square of the increase) turns out to be strictly valid only under ideal conditions. In practise, as the experiments of Andrews (1) have shown, serious and unsuspected errors may arise, due to varying degrees of reflection of sound waves from the floor, walls and other solid objects in the room. It follows that all tests of auditory acuity made indoors and in which the attempt is made to grade acuity in terms of distance are open to objection if anything like careful work is to be done.

The most obvious way to avoid this error is to use the *method of constant range, i. e.*, to station *E* and *S* always at the same place, to keep the position of all objects that might reflect the sound also constant, and to measure acuity in terms of the percentage of errors made in a series of tests at this selected range. For simple schoolroom diagnosis the stations of *E* and *S* may then be so chosen that all *S*'s whose acuity is known to be normal can just hear the sound used for the test.

B. Sources of Sound. (1) The primary advantage of *speech tests* is that they measure directly the most important function of the ear—the hearing of conversational speech, whereas all instrumental tests, because they test the perception of only a limited number of auditory qualities, fail to give unequivocal indication of auditory efficiency. One may hear the watch at a considerable distance and yet be relatively deaf for speech, or conversely. Speech tests should, accordingly, be given the preference where possible.

The use of speech tests is, however, rendered difficult for several reasons. (*a*) Speech involves a great variety and complex combination of pitches of varied intensity and clang-color, and these elements are further varied by changes in accent, emphasis, and inflection. To render speech tests available, therefore most careful study must be made of the elements of spoken and whispered speech, and lists of test-words must be prepared in the light of this analysis.[1]

(*b*) Examiners cannot guarantee uniformity of enunciation and intensity of stress in conducting the test, so that the results of different *E*'s, or even of the same *E* at different times, are rendered difficult of comparison. This difficulty must be met both by preliminary practise and care on *E*'s part, and by ranking *S*'s relatively, in terms of the empirically determined norms for each particular test.

(*c*) The acoustic properties of the room in which the test is held markedly affect its outcome. The method of relative ranking, coupled with the method of constant range, may be used to meet this difficulty.

(*d*) Unavoidable noises are more likely to interfere with speech tests than with tests conducted at close range, *e. g.*, by the audiometer. To offset this, tests must be conducted in as quiet a room as possible, and doubtful cases must be retested under the most favorable conditions that can be secured.

---

[1]This work has been done by Wolf (21). English number-word lists have been prepared and tested by Andrews (1). Reference to these writers will make clear why disparate words form the best speech-test material, and why numbers form the best type of words. Politzer's objection to numbers (12, p. 117) is answered by Bezold (2, p. 5; 3, p. 206).

Limits of space will usually preclude the use of vocalized or conversational speech, but *whispered speech* may be used for tests in a range of from 17 to 40 m., or about one-third that of vocalized speech. Whispering reduces the intensity of the vowels, whereas consonants are little changed. This test serves perfectly well for the practical examination of hearing and should be employed whenever feasible.

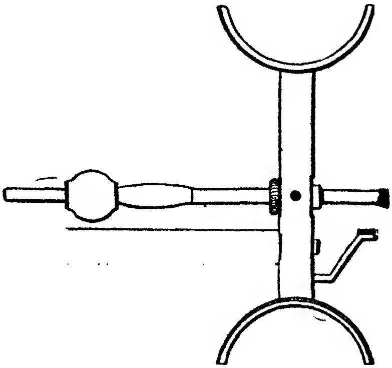


FIG. 48 POLITZER'S ACOUMETER.
From Titchener, *Experimental Psychology.*

(2) Reverting to *instrumental tests,* we find that the *watch* is most widely used. Its advantages are its convenience and accessibility and its relatively short range. Its disadvantages are that, like any instrument, it fails adequately to test the capacity to perceive speech, that its ticking is so familiar that illusions of hearing are frequent, and that watches vary in the intensity and quality of their ticks.[1] Proper attention to method will, however, reduce these disadvantages appreciably.

Various forms of *acoumeter* have been invented to meet the deficiencies of the watch. The instrument invented by Politzer (Fig. 48) is best

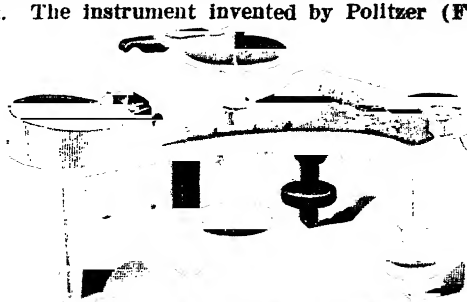


FIG. 49. LEHMANN'S ACOUMETER.
(Improved by Titchener.)

[1]Statements sometimes made in books on hygiene that, if the ticking of a watch can be heard at so-and-so many inches, the subject has normal hearing, are obviously absurd. The normal range for a watch-tick is given at 2.5 to 4.5 m., but one in the author's possession has a range of 12 m. See Bezold (3) and Sanford (14, p. 55.)

known and is extensively employed in clinical work. Its range is commonly given at 15 m., but will vary one or two meters from this, as test conditions vary. This acoumeter yields a brief tone, 512 vd., of constant intensity. It is usually employed with the method of varied range. For description, see Politzer (12, pp. 107-8). The upright is held between the thumb and forefinger, and the small hammer is dropped upon the steel cylinder from a constant height. A small disk attached to a pin, not shown in the cut, is used for bone-conduction and other diagnostic tests.

Lehmann's acoumeter (Fig. 49) has the advantage of allowing variation of intensity, and is thus better adapted to the space limits of the ordinary laboratory,[1] but it is, nevertheless, not very satisfactory for practical testing.

The acoumeter described by Toulouse, Vaschide, and Piéron (18) substitutes a drop of distilled water for the metallic ball, and an aluminum disk for the receiving plate.

Many attempts have been made to devise an instrument that will permit testing at the ear itself, in order the better to rule out disturbing noises. Commonly, these devices are electrical in nature, and are planned to utilize a telephone receiver in which clicks or tones are produced in a graded series of intensities.[2] Typical of these instruments is Seashore's

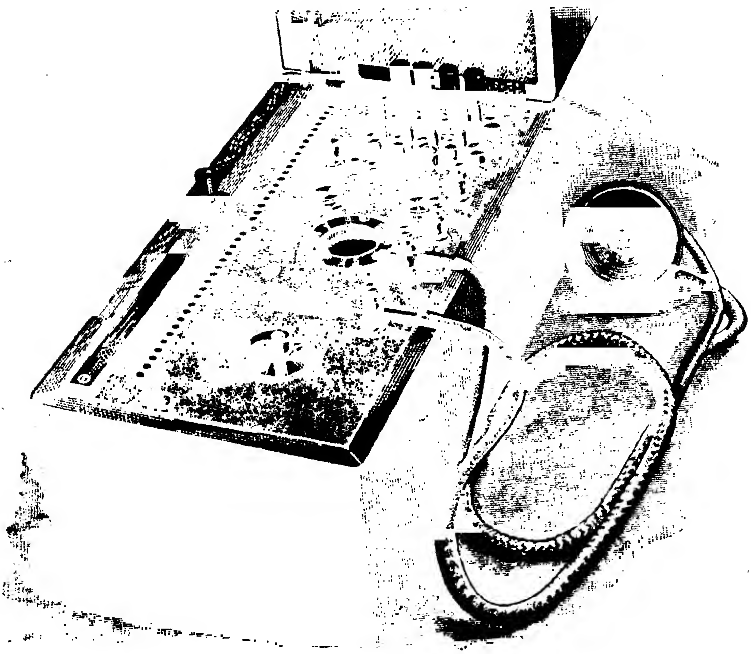


FIG. 50. SEASHORE'S AUDIOMETER.

[1]For description, see Hansen and Lehmann (7).
[2]For an extended discussion of the technique, and particularly of the calibration of this type of apparatus, consult Bruner (5).

audiometer (Fig. 50), which has been fully described by its inventor (15), and which has been extensively employed by him (16) and by others, *e. g.*, by the Child Study Bureau at Chicago (9, 17). Where its cost is not an obstacle this audiometer is to be recommended as the best device for testing hearing by the use of faint clicks or tones.

The audiometer invented by McCallie is nearly as expensive as the Seashore instrument, and is very far from being as satisfactory. It is open to objection in principle and in construction, and the author's experience with it does not confirm the favorable endorsements passed by some investigators.

Tuning forks may be employed for acuity tests in accordance with the method first suggested by Von Conta (20), in which a 512 vs. fork is struck and brought before the ear to be tested, and acuity determined by the length of time it can be heard. Blake's fork (Fig. 51) is devised especially for use by such a method, and may also be used for simple diagnostic tests as described in detail below.

A. WHISPERED SPEECH TEST

MATERIALS.—Meter stick. Telegraph snapper, for signalling. A number of small rubber stoppers, for ear plugs. List of 100 test-numbers arranged in ten series, as in the following Table.

TABLE 27.

Test-numbers for Auditory Acuity (Andrews)

I	II	III	IV	V	VI	VII	VIII	IX	X
6	84	19	90	25	14	8	52	73	24
29	69	53	7	13	31	93	35	41	95
42	17	34	39	46	9	27	64	16	62
87	92	28	62	7	65	60	81	95	49
53	33	97	84	54	98	15	6	57	80
94	26	45	21	70	76	74	19	38	71
70	50	72	56	91	40	36	78	20	16
35	75	60	75	83	23	49	40	89	3
18	48	3	43	68	52	82	23	64	58
61	1	86	18	92	87	51	97	2	37

PRELIMINARIES.—Select, if possible, an oblong room of average proportions and a length of at least 30 m. By rough preliminary tests, establish a range in this room such that not over 90 of 100 test-words can be correctly heard by a normal ear. If space will not permit this range to be established otherwise, interpose screens between *E* and *S*, or place *E* and *S* in adjoining rooms, off a straight line. The range may thus be cut down to from 18 to 20 m., or even less. Whatever may be the arrangement that affords a suitable range, make careful note of all acoustic conditions, *e.g.*, distance of range from walls, dimensions of rooms, exact position of *E* and *S*, disposition of large pieces of furniture in the rooms, number of doors

or windows opened or closed, time of day, etc. Be sure always to work under precisely these conditions.

METHOD.—(1) Seat *S* at the end of the range selected, with his right ear toward *E*. Carefully close the left ear by means of a rubber stopper inserted into the meatus.

This must completely close the ear, but must not be distressingly tight. *E* should practise on himself beforehand. If both ears are properly stopped, the ticking of a fairly loud clock can be heard only with difficulty when 1 or 2 m. away, and an ordinary watch cannot be heard when held close to the ears. The plug of cotton often used is entirely inadequate. Inserting the moistened finger-tip into the meatus makes an effective plug, but the position is uncomfortable, and *S* is likely to move the finger and thus to cause distracting noises in the stopped ear. The same objection may be made to the practise of stopping the ear by pressing in the tragus, or by closing the meatus with the fleshy part of the ball of the thumb.

Direct *S* to close or shield his eyes during the test, and on no account to watch *E*'s lips. His mouth must likewise be closed, since hearing is altered when the mouth is opened.

Give *S* a short preliminary series without recording results, until satisfied that he understands the conditions of the test and feels at ease.

(2) For the more formal test, pronounce the 100 words (or but 50, if time is limited) in groups of 10, in the following manner: at the conclusion of one expiration, snap the sounder once as a ready signal for *S*: at the conclusion of the next expiration, pronounce the test-number in whispered speech with the residual air in the lungs: then snap the sounder twice to indicate that the word has been pronounced, and let *S* either speak or write down the number that he has heard (using a dash if nothing is heard). Meantime, *E* interpolates three complete breaths, then gives the warning signal, then the test-number after the fourth breath, and so on until 10 test-numbers are given. After a brief rest, try the second 10 numbers, and similarly, the third, fourth, etc. To avoid possible error, let *S*, if he is writing his report, begin a new column with each ten.

(3) Stop *S*'s right ear and test his left ear in the same manner.

(4) Test *S*'s binaural hearing by letting him *face E*, but with precaution that he does not secure visual aid from *E*'s lips. This test is important, because binaural hearing may not be related to monaural range, and it is the type of hearing actually

used in daily life. If time is very restricted, test this form of hearing alone.

TREATMENT OF RESULTS.—*S*'s acuity is determined by the percentage of test-numbers correctly heard, in relation to the normal percentage which has been ascertained by averaging the percentages of all *S*'s tested under the same conditions. Thus, if the normal percentage be 70, and *S*'s be 60, his acuity is 6/7; if *S*'s be 80, his acuity is 8/7, *i.e.*, supra-normal. Credit may be allowed for partially correct reports, *e.g.*, 62 for 65: such allowance is specially recommended if 50 or fewer test-numbers are used. If *S* can hear nothing at all at the standard range, *E* may secure an approximate idea of the degree of defectiveness either by shifting to vocalized speech and then raising the intensity of the stimulus words until *S* hears correctly or by continuing the whispered speech at lesser distances (method of varied range).

NOTES.—The sounder is used to avoid changing *E*'s vocal 'set.'

If during the test, *S* becomes restless or inattentive, defer its completion.

It is best to test but one *S* at a time: two *S*'s may, however, be placed back to back, for testing the right ear of the one and the left of the other, if precaution is taken to ensure against communication or disturbance. If the room is large, and preliminary tests warrant the belief that acoustic conditions will be identical, more *S*'s may be tested by seating them on an arc equidistant from *E*.

A very crude group test may be carried out by placing all the children in a room at the limit of the ordinary classroom distance. Let them all close their eyes; then order them in a whisper to perform some unusual movement, such as to place the right forefinger on the palm of the left hand. Repeat with similar commands. Note any children who fail to respond, or who do so in evident imitation of others. Give these more careful tests later. Or take smaller groups of 10, similarly placed across the classroom. Provide each with a block of paper and pencil. Try a series of 10 whispered number-words, and let each write them as heard. Test carefully any who make a single error.

B. THE WATCH TEST

APPARATUS.—Ordinary watch, preferably *E*'s own watch or one that will be available for all the experiments. Meter stick. A number of small rubber stoppers for ear plugs.

PRELIMINARIES.—Select a room with a straight range of at least 10 m. Chalk off upon the floor a range of 9 or 10 m. subdivided into half-meters. (Use the meter-stick itself, held at right angles to the ear, for measuring distances less than 1 m.) Seat *S* at one end of the range thus marked off, preferably in a position that will not bring him within 2 m. of a reflecting wall or other surface, with his unused ear plugged and eyes and mouth closed, as in the speech test. Remove his watch if he carries one.

METHOD.—Hold the watch in the palm of the right hand with its face clearly exposed toward *S*, and on a level with his ear.[1] Start close to *S* and move out along the range-line, testing at each half-meter, until a point is reached at which the ticking is no longer heard. Do not exact continuous attention to the sound, but instruct *S* to reply "yes" or "no" when asked if he hears the watch. Give a signal "now" at each point where his judgment is to be made, but at every second or third "now," cover the watch with the left hand so that *S* shall not consciously or unconsciously report "yes" when he should report "no."

Reverse the procedure by approaching from a point certainly beyond *S*'s capacity to a point where the watch can certainly be heard, again by half-meter steps.

Repeat these two determinations until the results for each are consistent and average the two distances (moving outward and moving inward) to secure the range for the ear tested.

Test the other ear in the same manner, or both ears simultaneously, as directed in the speech test, if time permits but one determination.

---

[1]In this position the palm of the hand acts as a sound reflector, and any variation in the location of the watch within the hand must be avoided if the intensity of the sound thrown out is to be kept uniform. Thus, if the watch be held by its stem, free from the palm, there will result an astonishing reduction of the sound directed toward *S*'s ear.

Treatment of Data.—The normal range for the particular watch and under the particular test conditions must be determined *empirically* by collecting data from a number of *S*'s. It will then be found that the distance the watch is heard will be approximately the same for the majority of *S*'s, that a few will show exceptionally keen acuity, while *S*'s with defective hearing will range all the way from those just under the modal capacity to those who are unable to hear the watch when it is applied directly to the ear. While it is impossible to lay down rules, it may be said that with the ordinary watch having a normal range of 3-5 m., a reduction to 1 m. would signify a distinct deficiency in hearing.

C. TUNING FORK METHOD

Apparatus.—Blake's fork (Fig. 51). Stop-watch. [Rubber tube.]

Method.—(1) Stand directly behind *S*. Sound the fork by pressing the tips of its prongs together until they touch, and then suddenly releasing them. Hold it horizontally, with its plane of vibration vertical, opposite, and close to the ear to

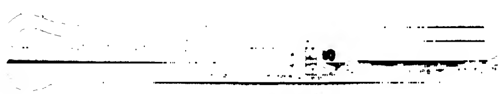


FIG. 51. BLAKE'S FORK.
For acuity and diagnostic tests by the temporal or 'ringing-off' method.

be tested. Lift the prongs away from the ear occasionally, so that *S* can state more easily when it actually ceases to be heard. Record the time by means of the stop-watch. Repeat 5 times with each ear, or until accordant times are given. Compare this time with the norm previously established empirically for the fork in use.

(2) For a simple diagnostic test, place the stem of the sounding fork between *S*'s teeth. If both ears are normal, *S* will hear the tone with equal intensity in each ear, or the tone may be subjectively located in the middle of the head. If, however, one ear is defective, the tone may be heard either more loudly or less loudly in the affected ear. If the tone is heard more

loudly in the ear which previous tests have shown to be defective, we may expect that the location of the defect on that side is in the middle or external ear, and that it may yield to proper medical treatment. If, on the contrary, the tone is heard better in the good ear, we may expect that the defect on the other side lies in the internal ear, or in more deeply seated portions of the auditory mechanism, and that it will probably not yield to treatment.

GENERAL RESULTS AND CONCLUSIONS.—(1) It is difficult to state the prevalence of defective hearing in school children, because of the arbitrary and loose nature of the tests that have been used, and the varying standards that have been set for normality of hearing. Thus, in New York City, a recent report indicates only 1.1 per cent. defective hearing; but here the test consisted merely in the use of a few whispered words in the school room at *20 feet* distance. The extensive Chicago tests,[1] conducted with Seashore's audiometer upon 6729 children, show that, if a pupil is classed as defective when the audiometer record is four points or more below the norm (indicating a defect such that "he would be seriously inconvenienced in detecting sounds of medium intensity"), 1080, or 16 per cent. of the numbers were defective in one or both ears (6.64 per cent. in both, and 9.55 per cent. in one year). A defect equivalent to three or more points of the audiometer scale was found in one ear in 26.3 per cent. and in both ears in 12.3 per cent. of those examined.

Other examinations are summarized by Young (22) as follows: "Sexton, of New York, examined 575 school children, of which 13 per cent. were hard of hearing; W. von Reichard, testing with the watch 1055 pupils of the gymnasium of Riga, found 22.2 per cent. with defective hearing. Weil, of Stuttgart, tested the sense of hearing in 5905 scholars of various kinds of schools, and found it below the normal in from 10 to 30 per cent. of the children, according to their social condition. Moure, of Bordeaux, found 17 per cent.; Gellé, of Paris, 22 to 25 per cent.; Bezold, of Munich, 25.8 per cent. of pupils with hardness of hearing." See also Chrisman (6) for a summary of investigations prior to 1893.

(2) With regard to the *partially deaf,* Macmillan and Bruner (9) conclude that, in theory, there exist varying degrees of deafness, "ranging all the way from slight and tempo-

[1]See Smedley (17), Macmillan (8, also summarized in 13).

rary impairment of hearing due to a cold, to the stage of absolute and permanent silence." An examination of the children attending the public day-schools for the deaf in Chicago, however, showed a somewhat unequal division of these pupils into 5 classes, based upon the somewhat conventional and immediately practical test of the status of the pupil in hearing in his schoolwork. Thus, of 174 cases, 55 were classed as totally deaf, 33 as "practically deaf" (hearing only intense and continuous sounds), 53 as possessing "a degree of hearing power" (hearing loud sounds, but not understanding vocal speech),[1] 25 as "deaf for ordinary school conditions" (hearing only words spoken loudly and close to the ear), and 8 as "hearing children temporarily needing special training in articulation."

(3) *Differences between the two ears.* Seashore (16) found decided differences in the acuity of the two ears, differences that were unknown to the *S*'s that exhibited them. Preyer, Fechner, and Bezold have concluded that the left ear tends to be the more acute: Bruner (5), however, as well as Miss Nelson (10), states that in both sexes the right ear is the more acute. Van Biervliet (19) asserts that inequality of hearing of the two ears is a universal fact, that the disparity is such that the poorer ear has a capacity 1/6 less than the better ear, but that the right ear is the better in right-handed, the left in left-handed *S*'s.

For practical purposes in connection with schoolroom tests, the determination of this difference is significant only when the inferiority of one ear is marked; in such cases, pupils should be so seated in the classroom as to bring their 'good' ear toward the teacher.

(4) Seashore's tests (16) indicate that acuity improves with *age* up to 12 years: this improvement is due partly to the development of the ear, but is slightly affected by the growth in ability to understand and to undertake the test.

(5) There are no noticeable *sex* differences, according to Seashore. Lombroso concludes that men's hearing is keener than women's.

[1]This class offers hope of improvement in hearing by means of mechanical devices for the intensification of speech.

(6) Seashore says there is "no indication that the *bright children* hear better than the dull children: there may be cases of children who are dull or are accounted dull because they do not hear well, but such cases are not common enough to be revealed clearly by our method, although there may be some indication of them." Nearly every other investigator, however, has found evidence to show that defective hearing has a positively injurious effect upon school-standing.

At Chicago (17) the examination of 5706 pupils with Seashore's audiometer showed that pupils below grade have, at every age, more cases of defect than those at and above grade, and that pupils in the school for backward and troublesome boys have a greater percentage of defect than boys of the same age in other schools. At Copenhagen, Schmiegelow found that, of 79 pupils regarded by the teachers as poorly endowed mentally, 65 per cent. had defective hearing. Similarly, Gellé found 75 per cent. of the defect in the pupils classed as poorest. Permewan, at Liverpool, averaged the distance the watch could be heard by 203 pupils when divided into three groups—bright, average, and dull—and obtained the figures 51 inches, 47.3 inches, and 31.25 inches for these three groups, respectively. Shermunski, at St. Petersburg, by means of the whisper test, found that, among those of normal hearing the ratio of good to poor students was 4.19 to 1; among those whose hearing was but ½ to ⅓ normal, the ratio was 2.6 to 1; among those whose acuity was less than ⅓, the ratio was 1.7 to 1.

(7) *Racial differences.* Bruner's St. Louis Exposition tests (5) indicated that the whites were clearly superior in acuity to the other races tested. The Filipinos had the poorest hearing of those tested.

(8) The simplest disturbance of hearing, if allowed to continue, may lead to serious results. In general, those who test the hearing of school children should note the condition of the ear, as well as test its capacity. Discharge of matter from the ear should be a cause for reference to medical attention.

(9) Children who are partially deaf should be guided, in their adoption of occupation, to avoid callings for which they are unfitted, *e.g.*, medicine, law, music, school-teaching, stenography, telephone or telegraph work, railroad, marine or military service.

(10) The ears of school children should be tested carefully at least once in two years.

(11) Defective hearing, like defective vision, may exist in serious degree and yet pass unnoticed by child, teacher, par-

ents, or friends. Of the 13 per cent. found defective by Sexton, only 3 per cent. were themselves aware of any defect, and only one of them was known to be deaf by his teachers.

NOTES.—In testing the hearing of those who are known to be partially deaf, *e.g.*, such a group as is mentioned in (2) above, the ordinary speech or instrumental tests are not serviceable. Use may, however, be made of the telegraph snapper mentioned in the first method, or of Blake's fork in conjunction with a 'differential tube.'

The noise of the snapper can be heard by the average ear at a distance of some 150 m. or more. In testing the partially deaf *S*, it should be held slightly behind his ear, out of direct view, and employed like the Politzer acoumeter, *i.e.*, by asking *S* to give the number of 'clicks' (2 to 5) that he hears. In very young *S*'s, sufficient indication of hearing may be obtained by watching for reflex starts of the whole body, or of some part of it.

The differential tube, as used by Macmillan and Bruner (9) consists of a tube of soft rubber 100 cm. long, and 4 mm. internal diameter, fitted with hard rubber tips for insertion, one into *S*'s, and one into *E*'s ear. After *S* has been familiarized with the sound of the fork by hearing it with the base applied to his front teeth, his ears are tested one at a time by placing the stem of the sounding fork upon the tube. On account, presumably, of the longer duration of the sound, this device may be used to detect a grade of hearing even lower than that detected by the snapper.

REFERENCES

(1) B. R. Andrews, Auditory tests. *AmJPs*, 15: 1904, 14-56, and 16: 1905, 302-326.

(2) F. Bezold, Schuluntersuchungen über d. kindliche Gehörorgan. Wiesbaden, 1885.

(3) F. Bezold, Funktionelle Prüfung des menschlichen Gehörorgans. 1897.

(4) W. Bingham, The rôle of the tympanic mechanism in audition. *PsR*, 14: 1907, 229-243.

(5) F. G. Bruner, The hearing of primitive peoples: an experimental study of the auditory acuity and the upper limit of hearing of whites, Indians, Filipinos, Ainu and African pigmies, N. Y., 1908. (Reprinted from *ArPs(E)*, No. 11.) See especially 55-108.

(6) O. Chrisman, The hearing of children. *PdSe*, 2: 1893, 397-441.

(7) F. Hansen und A. Lehmann, Ueber unwillkürliches Flüstern. Eine kritische u. exp. Untersuchung der sogenannten Gedankenübertragung, *PhSd*, 11: 1895, 471-530, especially 494 ff.

(8) D. Macmillan, Some results of hearing-tests of Chicago school children. *Medicine*, April, 1902.
(9) D. Macmillan (and F. G. Bruner), A special report of the Dept. of Child-study and Pedagogic Investigation on children attending the public day-school for the deaf in Chicago. Chicago, 1908. Pp. 88.
(10) Mabel L. Nelson, The difference between men and women in the recognition of color and the perception of sound. *PsR*, 12: 1905, 271-286, especially 280 ff.
(11) W. B. Pillsbury, Methods for the determination of the intensity of sound. *PsMon*, 13: 1910, No. 53, being pp. 5-20 of Rep.Com.Amer. Psych.Assoc. on the Standardizing of Procedure in Exptl. Tests.
(12) A. Politzer, Lehrbuch d. Ohrenheilkunde. Stuttgart, 1893.
(13) Rep. com. on statistics of defective sight and hearing of public school children. *RepComEd*, 1902, ii, 2143-2155.
(14) E. C. Sanford, A course in experimental psychology. Boston, 1895 and 1898. Pp. 449.
(15) C. E. Seashore, An audiometer. *UnIowaSdPs*, 2: 1899, 158-163.
(16) C. E. Seashore, Hearing ability and discriminative sensibility for pitch. *UnIowaSdPs*, 2: 1899, 55-64.
(17) F. W. Smedley, Rept. dept. child-study and pedagogic investigation. 46th An. Rept. Brd. Educ. Chicago, 1899-1900. Also *RepComEd*, 1902, i, 1095-1115.
(18) E. Toulouse, N. Vaschide and H. Piéron, Technique de psych. expérimentale, Paris, 1904. Pp. 330.
(19) J. van Biervliet, L'asymétrie sensorielle. *BuAcRoySci*, 34: Série 3, 1897, 326-366.
(20) Von Conta, Ein neuer Hörmesser. *Archiv. f. Ohrenheilkunde*, 1: 1864, 107-111.
(21) O. Wolf, Sprache und Ohr, Braunschweig, 1871. Also various articles in otological journals, as *Arch. f. Augen u. Ohrenheilkunde*, 3: Abth. 2, 35, and Abth. 1, 125; and *Zeits. f. Ohrenheilkunde*, vol. 20.
(22) A. G. Young, School hygiene. 7th An. Rept. State Brd. Health of Maine. Augusta, 1892. Pp. 399.

TEST 19

Discrimination of pitch.—Like other forms of sensory discrimination, this has been employed to discover the relation between such sensitivity and general intelligence. It has sometimes been employed to estimate musical ability, and it has, of course, general psychological interest. With adults and with children over eight or nine years of age, the test is relatively easy to administer.

APPARATUS.—Set of 11 forks —one standard fork of 435 vd. (double vibrations), and 10 comparison forks, whose rates are 0.5, 1, 2, 3, 5, 8, 12, 17, 23, and 30 vd. above the standard. A resonance box on which the forks may be mounted as they are used (Fig. 52). Soft-tipped hammer for striking the forks.

METHOD.—(*a*) *Preliminary trials.* Seat *S* with his back to the table at which *E* works, and about 1 m. distant. Instruct

him as follows: "When I say 'now,' close your eyes and listen carefully to the two tones you will hear; then tell me whether the second tone is higher or lower than the first. Say 'higher' if the second tone seems pitched above the first, 'lower' if below."

On the relative merits of reeds, blown bottles, vibrating strings (like the sonometer) and tuning forks for use in testing pitch discrimination, consult Seashore (8). Some of the results frequently quoted for discrimination of pitch by school children have been procured with the aid of instruments of very questionable accuracy. Thus Gilbert's tone-tester (3), which is constructed from an adjustable reed pitch-pipe, varies as much as five vd. in pitch with variation in the force with which it is blown. Stern's blown-bottles or tone-variators (10) necessitate a constant air-supply, and even then do not yield pitches which correspond to the attached scales. The sonometer (monochord or dichord) employed by Wissler (16) and Spearman (9) is rather unwieldy, not always constant in pitch and tone-color, and complicated by certain mechanical difficulties, while its pitches must be computed at each test in order to guarantee correctness of the assigned vibration-rate values. The instrument is defended, however, by Spearman (9, 243f). Wissler's method of using the monochord, in accordance with which *S* was obliged to manipulate the instrument, is indefensible, and his results are worthless, as far as pitch discrimination is concerned. The use of tuning forks in which the pitch of the comparison fork is varied by weights or riders (for illustration, see Titchener, 13, i, 68) also necessitates the computation of the pitch differences by counting beats, and both this and the manipulation of the riders is not easy for inexperienced *E*'s. For these reasons, a series of carefully tuned forks, selected for uniformity of tone-color, one for each pitch desired, is here recommended. The present apparatus (Fig. 52) has been described fully by the author in conjunction with Titchener (14).

Seashore (8) objects to the wooden resonance box and to the striking of the forks while on or before a resonator. He recommends holding the fork in the hand, striking it upon a rubber-covered stick and then holding the vibrating prongs before a suspended metal resonator. I have not been able to follow his method without introducing more sources of error than those encountered by the method here proposed.

S's judgment must always be in terms of the second tone. To request him to answer merely "same" or "different," as some investigators, *e. g.*, Gilbert, have done, would produce different results, as it is commonly less difficult to judge a difference than to judge the direction of this difference. In this preliminary series, *S* may be allowed to give the answer "same," if he naturally does so when he is unable to say "higher" or "lower."

S's who are extremely unfamiliar with tones occasionally do not understand what is meant by 'higher' and 'lower,' and, like markedly unmusical or tonally-deaf *S*'s, are apt to search for differences in intensity or duration of the tones instead of for qualitative (pitch) differences. In such an event, *E* must select two forks that give the maximal difference, and give *S* a short course of training by striking the forks in succession, and explaining after each pair that the second was higher or lower, as the case may be. If this training is futile, *S*'s discrimination must be extremely poor, and he may be ranked 30+.

Insert firmly in the oak pedestal, with their axes at right angles to the main axis of the resonance box, two forks that afford a large stimulus difference, *e.g.*, the standard and the highest fork (marked 30). Damp one fork, *e. g.*, the nearer one,

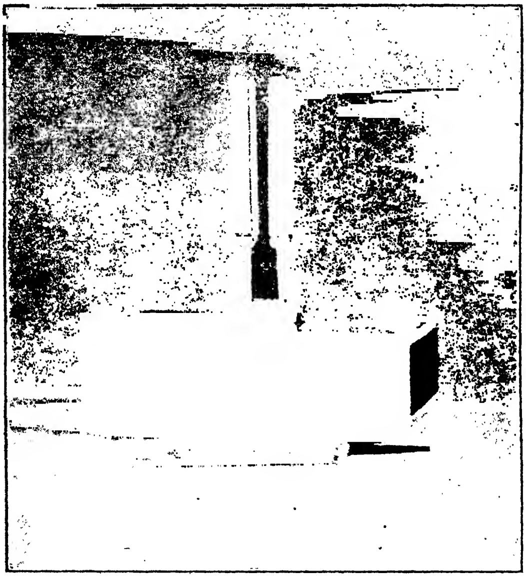


FIG. 52. TUNING FORKS FOR PITCH DISCRIMINATION.

by placing the left forefinger on the tip of one of its prongs: sound the other fork by striking one prong a clean light tap at a point about 1/4 the distance from its tip. Let the fork ring about 2 sec., then damp it by resting the middle finger upon it. After an interval of 2 sec., lift the forefinger, sound the second fork (while the 1st is still damped) and damp it similarly at the end of the 2 sec. Keep these time relations—2 sec. 1st tone, 2 sec. interval, 2 sec. 2d tone—constant, and strike the forks as uniformly as possible in all tests.[1] *S* always judges in terms of the second tone.

[1]Seashore (8) prescribes 1 sec. for the duration of each tone and 1 sec. for the interval. The author does not find the argument for these durations conclusive, and has obtained better results in practise from the use of the longer periods. Seashore's insistence of a tone of very mild intensity—just loud enough to be clearly heard by *S*—is fully justified. *E* should hold the hammer only 4-5 cm. from the forks and strike them a light tap with the thick felt tip of the hammer.

Continue the practise series, in accordance with the general plan for discrimination work described in the opening pages of this chapter, by inserting other comparison forks in place of the "30" fork, giving sometimes the standard, sometimes the comparison stimulus first. This preliminary series is to familiarize *S* with the general nature of the test, and to afford *E* a rough notion of the limits of *S*'s discriminative capacity. For most *S*'s, at least 4 or 5 min. should be given to this practise.

(*b*) *Test proper.* If *S*'s 'critical region' is not yet evident, give a formal series of pairs of stimuli, beginning with a supraliminal difference and passing toward subjective equality, until a difference is reached which *S* mistakes, or recognizes with difficulty. When, through this procedure or through the practise series, two forks are found that seem to be just barely different to *S*, keep these forks on the resonance box, and give a series of ten pairs of stimuli—five with the standard first, five with the comparison first, but in an irregular order. In this series *S* must not be allowed to judge "same," but should be made to guess in case of doubt. He should know that two different forks are being used, but should not know the direction of the differences which he is judging. If he gives 10 correct judgments, select for the next series a comparison fork nearer the standard in pitch; if he gives but 5 or 6 correct answers, and these with difficulty, select a comparison fork farther from the standard. Seek a pair of forks which will yield about 8 right cases in 10. Confirm the difference limen thus secured by trying series with comparison forks just sharper and just flatter than the one in hand. *S*'s discriminative capacity for pitch is indicated by the difference in vibrations between the standard fork and the comparison fork that yielded 80 per cent right answers.

RESULTS AND CONCLUSIONS.—(1) The *difference limen* for highly practiced *S*'s in careful laboratory tests is, for this region of the tonal continuum, about 0.3 vd.[1]

(2) The *pattern distribution* undoubtedly follows a skewed curve like that shown in Fig. 53, which is reported by Seashore

[1]For a summary of the work of Delezenne, Seebeck, Preyer, Luft, Meyer and others, consult Titchener (13, Pt. ii., 235 ff).

(8), and represents the records of 781 college students, 296 men and 485 women, for the differences indicated on the abscissa only, and without previous special training. On account of the skewness of the distribution and the absence of any well-defined mode, it is difficult to state what is the average or the modal capacity, but it is evident that more than one-half of those tested had a discrimination of 3 vd. or less. Similarly, Norton's test of 276 college students furnished an average of 6 vd., M.V. 4.2, and mode 3.5 vd., while in two-thirds of the cases the discriminative limit lay between 0 and 4.5 vd.

(3) *Dependence on sex.* According to Seashore, "pitch discrimination does not vary in any constant manner with sex" (8, p. 56). This result is at variance with the conclusions of Wissler (16), Thompson (12) and Burt and Moore (2), all of whom found females superior to males. The results of Wissler and Thompson are less reliable than those of Seashore: those of Burt and Moore deserve more attention. These investigators tested English school children and found for the boys a median of 6.0 vd., for the girls a median of 4.9 vd. They compute that only 21 per cent. of boys have a discrimination exceeding the median discrimination of girls, whereas in adults about 40 per cent. of men have a discrimination exceeding the median of women. The 'sex-difference' in favor of males is, therefore, negative for this test, being—29 per cent. for children and —10 per cent. for adults.[1]

(4) *Dependence on age.* While the relation of pitch discrimination to age is not entirely clear, it would appear to be established that from the time when a child can undertake the test intelligently there is no very great improvement in discrimination with age. Certainly individual differences appear at an early stage: some very young children can discriminate 2 vd. and less with certainty,[2] while others remain virtually

[1]To avoid possible confusion it may be explained that the sex-difference, following Burt and Moore, is computed on the basis of the relation of male to female performance. Thus, in the above instance, the median of female performance is, of course, exceeded by 50 per cent. of women, so that male performance is 10 per cent. short of this standard.

[2]Seashore says that "in a bright child with a good ear the physiological limit can be established for all practical purposes as early as the age of five."

tone-deaf throughout life. Thus, the distribution displayed in Table 28[1] is not markedly different from that displayed in Fig. 53. The figures published by Gilbert, whose 'tone-tester' is, as has been remarked, a questionable device, show a minimum of capacity at 6, a rapid improvement to 9 and then a gradual improvement to 19, with seeming losses of capacity at 10 and 15, which are referred by Gilbert to pubertal and other disturbances.[2]

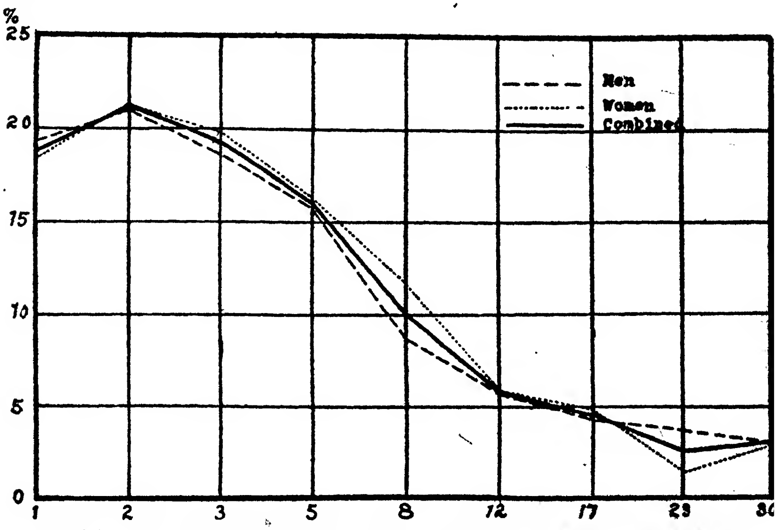


FIG. 53. THE DISTRIBUTION OF THE CAPACITY TO DISCRIMINATE PITCH (SEASHORE).

TABLE 28

Pitch Discrimination of 167 Children, Aged 6-15 Years (Seashore)

NO.	LIMEN IN VD.	NO.	LIMEN IN VD.
20	1 to 2	21	12 to 30
63	3 " 5	14	Over 30
48	6 " 10		

[1]These results are subject to the qualifications that the figures obtained from those of the children under 9 years of age are not very reliable. In some 30 cases not here included, Seashore was unable to determine a limen within the time at his disposal. Gilbert found only 3 of 190 children who could not discriminate a half-tone, *i. e.*, about 30 vibs. in this region, but as already stated, his *S*'s were asked to judge only difference.

[2]Cf. similar disturbances found by Bryan at 10 and 15 in motor tests.

(5) Seashore found a slight positive correlation between pitch discrimination and *auditory acuity* (Test 18).

(6) (*a*) *Practise* does undoubtedly improve pitch discrimination, but investigators are not in agreement as to the extent of such improvement. It seems evident that its limits are fixed by anatomical and physiological conditions in the ear itself, and that these limiting conditions vary in different individuals. In general, the improvement is not as great as that observed in some other functional capacities, *e.g.*, the discrimination of dual cutaneous impressions (Test 23), and is reached after so short a period of training as to justify the statement that pitch discrimination is an inborn capacity.

Seashore believes that maximal capacity can be attained after very little practise. In 20 days training, he found that some *S*'s exhibited no improvement, while the maximal improvement reported was the reduction of the limen of an unmusical *S* from 30 to 5 vd., and he believes that an ingenious *E* can discover "the proximate physiological threshold to a fair degree of certainty in a well-planned half-hour individual test."

Spearman (9, p. 231) believes that 15 minutes fore-exercise will reduce the pitch-limen by an amount depending very largely upon *S*'s previous general familiarity with tonal experiences. Thus, he computes a reduction for specially practised *S*'s from a limen of 0.5 (before the 'exercise') to 0.3 (after the special exercise), for musicians from 4 to 2 vd., for non-musicians of general culture from 10 to 4 vd., and for European villagers from 30 to 8 vd. These figures would indicate that even a practised musical *S* profits by a preliminary 'warming-up,' and they emphasize the importance of giving such fore-exercise to all *S*'s.

(*b*) Aside from practise in the narrow sense, *i.e.*, special exercise in pitch-discrimination under experimental conditions, we may consider the effect upon discrimination of practise in the wider sense, *i.e.*, of general *musical training*. Seashore is, again, very emphatic in his declaration that individual differences in pitch discrimination are not due principally to musical training, and Spearman's conclusion is that, "though a correspondence really does exist, yet it is not to the smallest degree of the specific character contemplated by those who talk of 'musical sensitivity,'" *i.e.*, by those who refer to pitch discrimination as a test of "musical sensitivity."

To the author, this seems a case of one-way correspondence; an individual who cannot discriminate a half-tone cannot be musical, but an individual who is not musical *may* have a perfectly good discrimination after a little preliminary practise. Given, however, a good natural

capacity for discrimination, it is unquestionably true that musical training tends to keep this capacity up to the individual's physiological limit. In testing 50 grammar-school boys for pitch discrimination, the three best discriminators were found to be "taking lessons" on the violin. The author has also shown elsewhere (15) that in the case of an unpractised, unmusical *S*, it may be possible to reduce the limen very decidedly by working under very constant, favorable conditions—such as duration, intensity, timbre of tones, time-intervals, etc.—but that the slightest modification of these conditions will make discrimination very difficult or impossible.

The fact that the test of pitch discrimination measures a specific capacity, apparently dependent upon the structure of the sense organ and to this extent unaffected by exercises for training, justifies the use of the test, according to Seashore (7 and 8), as a tentative classificatory device. The relation between the difference limen and the educational 'prognosis' of the child is set forth thus:

Below 2 vd.: May become a musician;

3-8 vd.: Should have a plain musical education (singing in school may be obligatory);

9-17 vd.: Should have a plain musical education only if special inclination for some kind of music is shown (singing in school should be optional);

18 vd. and above: Should have nothing to do with music.

(7) *Correlation with intelligence.* Seashore reported no correlation between pitch discrimination and intelligence, when general intelligence was indicated by class standing and teachers' estimates, and the correlation was worked out by the method of group-classification. Spearman terms this an "ingenious," but somewhat "disseminated" method, and by subjecting Seashore's results to his (Spearman's) methods, obtains from them a correlation index of 0.24 ± 0.07. From his own results, Spearman concludes that general intelligence is correlated with pitch discrimination by the index 0.94, or, as he states, "The [Intellectual] Function is 9 parts out of 10 responsible for success in such a simple act as Discrimination of Pitch."[1] Later, in conjunction with Krueger (4), Spearman computes a correlation of .83 between the capacity for pitch discrimination and the hypothetical "central factor," later identified with "General Ability." Another Englishman, Burt (1), has also reported results which tend to confirm the position of Spearman as against that of Seashore. Burt tested 320 normal children in elementary and preparatory schools and

[1]It is but fair to call attention to the fact that Spearman's formulas have been called in question.

obtained corrected coefficients of .52 and .41 for the correlation with the teachers' estimates of intelligence.[1]

The author's test of mental and physical ability in fifty 8th-grade boys included a determination of pitch discrimination, both with the Stern tone-variator and with a monochord. These two tests showed a correlation of .83. The variator test, which was, on the whole, most reliable, showed a correlation of .27 with class standing.

(8) *Other correlations.* Krueger and Spearman report 'raw' correlations between pitch discrimination and both adding and the Ebbinghaus completion test, of .67 and .59, respectively. After the application of the expanding formula (see Ch. 3), these correlations become .80 and .81, respectively. Thus it appears certain to these authors (4, p. 78) that the capacity to discriminate pitch actually exhibits a very high degree of correlation with the seemingly fundamentally different capacities requisite in adding and in the Ebbinghaus test.

Norton (5) compared discrimination of pitch with discrimination of intensity of noise and found a correlation of .39 (median ratio .67).

The author found a correlation of .27 between pitch discrimination and the discrimination of lifted weights.

NOTES.—With reference to musical ability, Stumpf (11, ii., p. 157) proposes as tests: (1) discrimination, (2) ability to sing a note struck on the piano, (3) ability to judge whether one or two tones are present in various fusions, (4) skill in determining the relative pleasantness or unpleasantness of two chords separated by a short pause. M. Meyer denies the validity of the discrimination test for musical ability, and favors a form of test in which *S* is asked to state whether a given bass note does, or does not, form the proper fundamental for a given chord (played in the treble region of the piano). Seashore regards the discrimination test as first and fundamental, and

[1]In commenting upon the relatively high correlation found by himself and by Spearman between discrimination of pitch and general intelligence, Burt thinks the relation is to be explained not as a manifestation "of a fundamental identity between Intelligence and General Sensory Discrimination, but rather historically, * * * by the large dependence of the development of intelligence in man upon power of speech, and of this, in turn, upon auditory acuity" (1, p. 132).

mentions as supplementary tests: "the sense of rhythm, and rhythmic action, tonal fusion (consonance and dissonance), auditory imagery, auditory memory, discrimination for intensity of sound, and vocal reproduction of a tone" (8, p. 58).

It is well to inquire of all *S*'s, before the test is administered, whether they are musical or not, whether they play any musical instrument, or sing, or are 'fond' of music. The author has found several instances of children who were quite unable to distinguish pitches several tones apart, but who were compelled by their teachers to take systematic instruction in singing along with other children in the public-school classes.

S's that fail to discriminate the 30 vd. difference may be further tested by a piano to see whether they are absolutely tone-deaf. If time permits, it is of interest to see whether *S*'s with very poor discrimination can be improved by systematic practise.

For best results, the discrimination test should be given individually; if necessary to undertake group tests, it is better to work with small groups of 5 or 6 *S*'s: supply them with pencil and paper; let them number the trials and write their judgments—"H" for higher or "L" for lower—after each number.

REFERENCES

(1) C. Burt, Experimental tests of general intelligence. *BrJPs*, 3: 1909, 94-177.

(2) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.

(3) J. A. Gilbert, Experiments on the musical sensitiveness of school children. *SdYalePsLab*, 1: 1893, 80-87.

(4) F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. *ZPs*, 44: 1907, 50-114.

(5) W. W. Norton, The correlation of pitch and intensity discrimination. *PsBu*, 7: 1910, 55-56.

(6) C. E. Seashore, Hearing-ability and discriminative sensibility for pitch. *UnIowaSdPs*, 2: 1899, 55-64.

(7) C. E. Seashore, Suggestions for tests on school children. *Ed. Rev.*, 22: 1901, 59-82.

(8) C. E. Seashore, The measurement of pitch discrimination: a preliminary report. *PsMon*, 13: 1910, No. 53. (Being pp. 21-60 of Rep.Com. Amer.Psych.Assoc. on the Standardizing of Procedure in Exptl. Tests.)

(9) C. Spearman, General intelligence objectively determined and measured. *AmJPs*, 15: 1904, 201-293.

(10) L. Stern, Der Tonvariator. *ZPs*, 30: 1902, 422-432.

(11) C. Stumpf, Tonpsychologie. 2 vols., Leipzig, 1883, 1890.

(12) Helen Thompson, The mental traits of sex. Chicago, 1903. Pp. 188.

(13) E. B. Titchener, Experimental psychology. Vol. II, Quantitative experiments. New York, 1905.
(14) E. B. Titchener and G. M. Whipple, Tuning forks for pitch discrimination. *AmJPs*, 20: 1909, 279-281.
(15) G. M. Whipple, Studies in pitch discrimination. *AmJPs*, 14: 1903, 553-573.
(16) C. Wissler, The correlation of mental and physical tests. *PsMon*, 3: No. 6, 1901. Pp. 62.

TEST 20

Discrimination of lifted weights.—We may compare two weights either by attending passively to the pressures set up when they are laid upon the skin, or by actively lifting or 'hefting' them. In the first instance, we see illustrated the procedure employed in testing discrimination of pressure (Test 21); in the second, that employed in testing discrimination of weight (in the narrow sense). The latter form of discrimination is always the keener, since to cutaneous pressure, there are added sensations contributed from muscle, joint, and tendon, particularly from the tendon. Because it is the movement of lifting the weight that aids us in estimating its comparative amount, the determination of this form of discriminative capacity is sometimes loosely termed the measurement of the 'muscle sense,' or the 'muscle sense test.'

Of the historical development of this important experiment, this is not the place to speak.[1] Attention is given here merely to the use of the experiment as a comparative test of mental or psychophysical efficiency. The most important investigations of this type are those of Burt (2), Gilbert (5), Thompson (10) and Spearman (8). The test was also included in those administered by the author to 50 8th-grade boys.

[1]The stock laboratory experiment is described and its technical aspects are discussed with sufficiently full citations of its literature by Titchener (11, Pt. I, 115 ff; Pt. II, 265 ff). As Titchener remarks: "This may be regarded as the classical experiment of quantitative psychology. On the psychophysical side, it has engaged a long line of investigators: Weber himself (13), Fechner (3) and Hering, all employed it to test the validity of Weber's Law; and a glance at the current magazines will show that the work begun by them has continued down to the present day. On the psychological side, it has been made by L. J. Martin and G. E. Müller (6) the vehicle of a qualitative analysis of the sensory judgment, the most elaborate and penetrating that we have." The recent work of Brown (1) should be cited in this connection.

Gilbert used weights of the 'cartridge' pattern, similar to those here prescribed. His method was less exact than could have been desired: his school children had simply to sort out all the weights that were the same as the standard, 82 g. As there were but 9 comparison weights, yielding a maximal weight of but 100 g., Gilbert encountered numerous cases (see Table 27) in which no discrimination could be made within this range.

Miss Thompson employed cartridge weights giving a range from 80 to 100 g., and apparently found no difficulty in testing adults with this equipment.

Spearman, similarly, employed cartridge weights with a standard of 1000 grains and with a series of geometrical increments, as proposed by Galton (4, Appendix). It is to be noted that the smallest increment of the original Galton series, 1/100, proved too coarse to test the capacity of some of Spearman's *S*'s, while the largest increment (mentioned by Galton for use with "morbid" cases) proved too fine to test the capacity of others.

A test carried on at Columbia University under the name "perception of weight" or "force of movement" consisted in lifting the handle of a spring dynamometer until it touched a stop. The 'reagent' then made 10 successive attempts to pull the handle to the same point when the stop was removed. It is evident that the results cited by Wissler (15) for this test are not comparable with those obtained by the standard form of the weight-discrimination test.

Van Biervliet (12) used weights of 500, 1000, 1500 and 2000 g. on the 'favored' side of the body in the case of 100 *S*'s that were tested by him for asymmetry in weight. The weights were lifted by a string attached to the index finger, and a simple gradation method was employed to determine the equivalent, for the left side, of a given weight on the right side of the body. The method is too unlike the standard method to admit of comparison of results. It may be stated, however, that the fraction 1/9, which the same author claims to have established as a constant of asymmetry in all sense-departments, was also found in this test, *e. g.*, a weight of 450 g. in the left was equal to a weight of 500 g. in the right hand, etc.

APPARATUS.—Blindfold or cardboard screen. Set of discrimination weights, comprising a standard, 80 g., and 23 comparison weights, yielding the series—80.5, 81, 81.5, 82, 82.5, 83, 83.5, 84, 84.5, 85, 86, 87, 88, 89, 90, 92, 94, 96, 98, 100, 105, 110, and 120 g. The weights are of identical size, shape, and color; are made of wood to avoid disturbing temperature sensations, and are marked inconspicuously (with reversed numbers), so that their weight may be known to *E*, but not to *S*.

METHOD.—Follow the general plan of procedure outlined in the introductory pages of this chapter. This plan embodies (1) a preliminary series of trials between the standard and various comparison weights to familiarize *S* with the conditions of the test and to indicate to *E* the probable 'critical region' in which *S*'s limit of capacity will be found, (2) a more formal determination of this region by systematic procedure

from too great to too small a stimulus-difference, (3) the selection from this region of a stimulus-difference (the standard and some single comparison weight) which may be expected to yield about 8 right judgments in 10, and which is given 10 times (5 times with the standard first, 5 with the comparison weight first—the arrangement being determined by chance), and (4) the final confirmation of the difference that yields 8 correct judgments in 10 by the trial of slightly smaller or slightly larger stimulus-differences, as may be required.

In the application of this procedure to weight discrimination, the following suggestions may be made. *S* should take his position, standing, before the table upon which *E* has arranged the weights. *S*'s view of the weights must be cut off, either by a well-arranged blindfold or by a horizontal cardboard screen so adjusted that he may lift the weights easily, but may not see them. In each trial, *E* selects a comparison weight, determines upon the order (standard first or comparison first); then, with a warning 'now,' places the first weight between *S*'s thumb and his first and second fingers: *S* hefts this weight, replaces it upon the table, when *E* quickly removes it and substitutes the second weight of the pair under trial, which is, in turn, hefted and replaced by *S*. The judgment must then be given promptly by *S* and always in terms of the second weight:—"heavier," "lighter," or "equal."[1] The details of the manner of lifting the weights may, in general, be left to each *S*. Fechner allowed 1 sec. for raising, 1 sec. for lowering, and 1 sec. for changing the weight, so that each comparison required 5 sec. for its execution.

TREATMENT OF DATA.—The difference between the comparison weight that yields 8 right judgments in 10 and the standard weight, 80 g., affords the absolute difference limen. The fraction formed by taking this difference as the numerator and the standard weight as the denominator affords the relative difference limen, and is the common index of efficiency in the test, since relative capacity is found to be nearly constant for a

[1]In the final test of 10 trials with a constant stimulus-difference, it is preferable to ask *S* to guess in case of an equal judgment; otherwise equal judgments may be recorded as wrong. *S* should not be permitted to return to the first weight after the second has been lifted.

given individual within a wide range of absolute weights.[1] It is, of course, the constancy of this fraction that constitutes the essential fact of Weber's Law.[2]

Results.—(1) *Normal capacity.* In one place, Weber cites as the average sensible discrimination for lifted weights for four *S's*, 3/32; in another place, he gives 1/40 as the difference just distinguishable by "quite the majority of human beings without any long preliminary practise" (13). Other authorities have placed the norm of performance at 1/17 or at 1/24 (Seashore, 7, p. 96). The author's tests with the apparatus and methods here described indicate for 8th-grade boys an average limen of 4.7 g. (standard 80 g.). The corresponding fraction, 1/17, is presumably close to the average performance for boys of this age. This result is corroborated by Spearman's series, which most nearly resemble the author's in method and apparatus; Spearman quotes 1/15 for a test made under unfavorable conditions and 1/20 for a test made upon older children under favorable conditions. The results reported by Burt for elementary school pupils (threshold 8.75 g., m.v., 1.5, extremes 6 and 16 g.) and for "high-class preparatory school" pupils (threshold 9.3 g., m.v., 1.6, extremes 5 and 11.5 g.), since they are based on a standard of 100 g., indicate a much less fine discrimination, possibly due to differences in the arrangement and execution of his tests.

(2) *Individual differences.* The work of every investigator has shown that the capacity to discriminate lifted weights differs very considerably among normal *S*'s, even when age, sex, and practise factors are eliminated. The author found 7 boys in 50 who could discriminate 80 and 81 grams, and one boy who could just discriminate 80 and 97 grams: reference has already been made to the fact that several experimenters

[1]Judgments of weight are to some extent affected by the height to which the weights are lifted. In case *S* shows distinct tendencies to lift the weights to unequal heights it may be necessary to arrange some device to regulate the amount of movement. Burt, for example, stretches a tape horizontally between two posts 17.5 cm. above the table.

[2]Recent work, however, raises a strong presumption of doubt as to the applicability of the law over all ranges. See especially Brown and Strong.

have found their weights inadequate to measure the wide differences in capacity that they encountered.

(3) *Dependence on age.* Spearman's tests convinced him "that the younger children were almost equal to the older ones and both were not far from adults," and also that there is no appreciable loss in weight discrimination with the coming of old age. Gilbert, however, as Table 29 shows, found a gradual improvement in discrimination from the 6th to the 13th year. Developmental disturbances appear from 12 to 14, and discrimination apparently does not improve thereafter. The method by which these results were obtained is, as already noted, open to criticism.

TABLE 29.

Dependence of the Discrimination of Lifted Weights on Age (Gilbert).

	6	7	8	9	10	11	12	13	14	15	16	17
Median Limen, Boys	13.0	13.2	12.2	10.2	8.6	10.2	7.6	6.0	5.0	6.2	6.0	6.0
Median Limen, Girls	16.8	13.2	11.0	10.0	9.2	7.6	7.6	5.6	7.2	7.2	6.8	6.4
Per cent over 18 g., Boys	26	36	35	23	12	5	0	5	0	0	2	0
Per cent over 18 g., Girls	49	40	28	17	12	6	6	0	0	0	2	2

(4) *Dependence on sex.* That boys are superior to girls and men to women in the discrimination of lifted weights is conceded by all investigators. This is indicated in most, though not in all of the groups of Table 29. Burt and Moore state (3, p. 366) that "the sex-differences in children and adults appear to be about + 40 per cent."[1] They also have computed the sex-difference discovered by Miss Thompson to amount to + 16 per cent. Miss Thompson is inclined to find the explanation for this difference in the same factors that make men superior in tests of motor ability, though this would not appear of necessity to affect a sensory capacity. Spearman, on the other hand, contends that "the fluctuating differences of sen-

[1]The difference, however, is given by these writers in their table, p. 369, as -48 per cent., despite the fact that the same table shows the median for boys to be 8.5 g., for girls 12.5 g. The correct excess percentage over 50 per cent. in favor of the boys is + 48. In other words, 98 per cent. of boys are better than the median girl.

sory discrimination observable in connection with sex at the various stages of growth are chiefly and perhaps altogether a mere consequence of similarly fluctuating differences in intelligence" (8, p. 261).

(5) *Dependence on practise.* The consensus of opinion is that, at least in comparison with many other mental activities, the discrimination of lifted weights is but little affected by practise. Thus, for instance, Biedermann and Loewit (quoted by Spearman) found that a difference limen of 1/21 fell only to 1/23 at the conclusion of a protracted research. It is also true that Fechner, who devoted most heroic amounts of time to weight discrimination, did not thereby attain remarkable capacity, and Brown, who made on a "faithful and devoted observer 75,100 experiments," concludes that there exists in these experiments no influence of practise. It is, nevertheless, possible that in the very early stages of this test, say in the first 15 min., there may be a decided improvement, especially for untrained and unskilled *S*'s. Spearman says that sometimes "the improvement is enormous." In general, then, the effect of practise upon this form of discrimination is very like that for the discrimination of pitch.

(6) *Correlation with intelligence.* Spearman's 'corrected' index of correlation between discrimination of lifted weights and general intelligence (rankings both for 'common-sense' and 'school-cleverness') for 24 village-school children amounted to .43, P.E. .10. In his work with high-class preparatory school pupils, under less desirable test conditions, the correlation was only .12, P.E. .09. Burt concludes from his own tests that the "connection between weight discrimination and intelligence seems to be either zero or even slightly inverse" (about —.10); and that "boys of superior cultural status are hardly as acute in distinguishing fine differences of weight as those of lower social status."[1] The author's tests of 50 school boys showed no correlation with intelligence.

[1]The discrepancy between their results causes Spearman to explain his higher correlation as due to a larger amount of practise, saying: "I have found over and over again that practise may greatly increase tendencies to correlate." The reader may compare the similar conclusion reached by Hollingworth, cited in Ch. 2, and the seemingly contradictory conclusion reached by Burt that practise in *successive* sittings tends to have the reverse effect on intelligence correlations, at least in tests whose procedure is readily grasped (2, pp. 123 and 168).

(7) *The feeble-minded.* Unpublished results of tests at Vineland, N. J., show that the feeble-minded fall into three groups with respect to the discrimination of weight—those who are totally unable to perform the test, those who succeed in discriminating by the use of simultaneous, but not of successive lifting, and those who succeed with successive lifting.[1] The results seem further to indicate that when discrimination can be made at all, the difference limen is not much different from that of normal *S*'s.

(8) *Other correlations.* The combined results for his elementary-school and preparatory-school pupils gave Burt the following values for correlations with the discrimination of lifted weights: esthesiometric limen .49, tapping .42, mirror drawing .30, pitch discrimination .29, the spot-pattern test .18, McDougall's dotting test .16, memory-span for words .14, alphabet test .10, comparison of lines .00, sorting cards —.05, dealing cards —.17. The coefficient of reliability for the discrimination test itself was .86 for the elementary and .51 for the preparatory school.

(9) Strong was unable to secure with the weight test indications of the defects clinically demonstrated in manic-depressive insanity.

(10) A *'constant error'* is exhibited by most *S*'s, in that they tend to overestimate the second weight. This, of course, makes it doubly imperative that the procedure be so arranged as to reverse the time-order in half the trials.[2]

REFERENCES

(1) W. Brown, The judgment of difference with special reference to the doctrine of the threshold in the case of lifted weights. Berkeley, 1910. Pp. 71.

(2) C. Burt, Experimental tests of general intelligence. *BrJPs*, 3: 1909, 94-177, especially 120-3.

(3) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 223-284, 355-388.

(4) G. T. Fechner, Elemente der Psychophysik, esply. vol. I. (Reprint, Leipzig, 1889, 1907.)

(5) F. Galton, Inquiries into human faculty. New York and London, 1883. (*Reprinted in Everyman's Library, No. 263.*)

[1] Attention may be called to the fact that in the Binet-Simon scale the test of arrangement of weights, which necessitates discrimination of lifted weights, is a 9-year test.

[2] On other qualitative aspects of this test, see especially Martin and Müller and Brown.

(6.) J. A. Gilbert, Researches on the mental and physical development of school children. *SdYalePsLab*, 2: 1894, 40-100.
(7) Lillie J. Martin and G. E. Müller, Zur Analyse der Unterschiedsempfindlichkeit. Leipzig. 1899.
(8) C. E. Seashore, Elementary experiments in psychology. New York, 1908. Pp. 218. See especially Ch. viii.
(9) C. Spearman, General intelligence objectively determined and measured. *AmJPs*, 15: 1904, 201-293.
(10) E. K. Strong, Jr., A comparison between experimental data and clinical results in manic-depressive insanity. *AmJPs*, 24: 1913, 66-98, especially 80-2.
(11) Helen B. Thompson, The mental traits of sex. Chicago, 1903.
(12) E. B. Titchener, Experimental psychology. Vol. II. Quantitative experiments. New York, 1905.
(13) J. J. van Biervliet, L'asymétrie sensorielle. *BuAcRoySci*, 34: Série iii., 1897, 326-366, especially 330.
(14) E. H. Weber, Der Tastsinn und das Gemeingefühl, in Wagner's Handwörterbuch d. Physiol., vol. 35, 1846. (Reprinted separately 1851.)
(15) F. L. Wells, On the variability of individual judgments. Essays in Philos. and Psych. in Honor of Wm. James. 1908.
(16) C. Wissler, The correlation of mental and physical tests. *PsMon*, 3: No. 6, 1901. Pp. 62.

TEST 21

Discrimination of pressure.—The determination of the difference limen for pressure, like that for lifted weights (Test 20), has constituted one of the standard psychophysical experiments since the time of E. H. Weber, who utilized it in connection with other tests to establish the well-known law that bears his name. By experimenting with standard weights of 32 oz. and 32 dr., respectively, Weber was able to report that "a difference of the smaller weights is not less accurately distinguished by touch than the same difference of the larger weights."

This test with 'resting weights,' sometimes termed the 'pressure sense' test, appears to have been less frequently used for functional and comparative purposes than the test with 'lifted weights.' Its feasibility depends very largely upon the type of apparatus employed. Differences between the temperature of the weights and that of the skin, variation in the temperature of the weights themselves, in the 'jar' of application, in the area and place of application, etc., must be excluded, since they inevitably produce conflicting results. To obviate these errors and to render the test more simple in execution and more reliable in outcome, the use of a 'pressure-balance,' following

the principle adopted by Merkel (7, p. 255) is desirable, if not essential.

Other forms of pressure-balance have been elaborated by Jastrow (5)—figured by Sanford (8, pp. 417-8) and by Titchener (10, pt. ii.)—and by Bolton and Withey (1). The balance here prescribed has been designed by the author (11) to supply in a single relatively simple apparatus a device for determining both the capacity for pressure discrimination and sensitivity to pain (Test 22). It may be regarded as a combination of the principle of Merkel's and of Jastrow's pressure-balances and that of Gilbert's balance-algometer (3).

APPARATUS.—The author's pressure-pain balance (Fig. 54). Cardboard screen with suitable supports. Seconds' pendulum (Fig. 28), or other noiseless device for controlling the time-relations of the test.

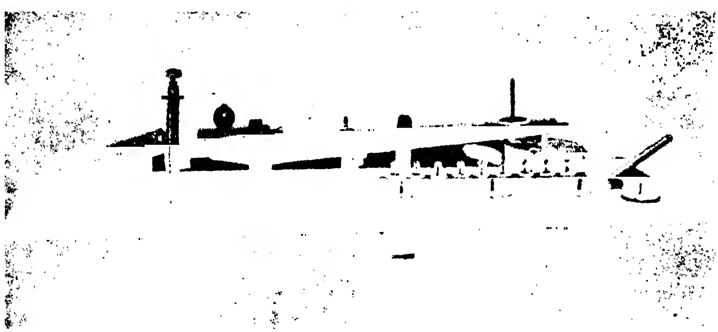


FIG. 54. PRESSURE-PAIN BALANCE.

PRELIMINARIES.—Place the balance upon a low table. See that the beam of the instrument moves freely, but comes to rest in a horizontal position when no weights are applied: if necessary, turn the small screw in the tip of the arm inward or outward until this position of rest is secured.

To assure comfort, the instrument should be so placed that *S*'s wrist will come just over the edge of the table; his elbow will not then be forced up into an awkward position, and his hand can lie upon the hand-rest, with the end of his forefinger projecting straight forward between the upper (stationary) and the lower (movable) tip of the balance. Adjust the upper tip so that it is in permanent contact with the center of the finger-nail, but does not touch the skin of the finger.

Arrange the screen to cut off *S*'s view of the apparatus, or, if he be reliable, simply instruct him to close his eyes.

Place the pendulum where its oscillations will be easily visible.

METHOD.—The general plan of procedure is identical with that outlined in the introductory pages of this chapter, and recapitulated in Test 20. To apply this procedure to the test with the pressure-balance, after throwing the lever down to the right, place the weight marked *B-100 g.* on the pin marked *B*, at the outer end of the beam. This weight is not removed during the experiment, and constitutes the standard stimulus. Place upon the second pin, marked *A*, the desired increment weight—any one, or any combination, of the weights marked *A*. To apply a pressure-stimulus, move the release-lever up to the left, so as to depress the support beneath the beam of the balance. To remove the stimulus, move the same lever to the right. The increment-weights are added to the standard stimulus when they rest upon the beam at *A:* they are subtracted from the total pressure, at will, by depressing the increment-weight lever, which lifts them from the beam and allows only the standard stimulus, 100 g., to be operative. Thus, for example, to test the discrimination of 150 g. and 100 g., move the release-lever down to the right, place upon the pin *A* the 30 g. and the 20 g. weights, and upon the pin *B* the 100 g. weight. Give *S* a warning "now," and 2 sec. later move the release-lever smoothly up to the left: allow the pressure (150 g.) to be felt for 2 sec., then move the release-lever to the right: immediately depress the increment-weight lever, and apply the second stimulus (100 g.) in the same manner, while this lever is held down. *S* judges, always in terms of the second stimulus, saying "heavier," "lighter," or "equal."[1]

The exact duration of the stimuli and of the interval between them is of less importance than constancy from trial to trial.

To avoid local fatigue, at least 15 sec. should elapse between successive judgments.

E must practise the manipulation of the instrument, and take particular precaution to move the release lever so as to

[1]Equal judgments, as previously explained, are to be avoided, if possible, in the final trials with a constant stimulus-difference.

avoid either too sudden application, which produces a disturbing 'bump' and vibration, or too slow application, which also renders the judgment more difficult.[1]

S must be specially instructed to receive the stimulus passively, so far as his finger is concerned. A downward movement of reaction in the finger tip converts the test, virtually, into a test of discrimination of lifted weights.[2]

TREATMENT OF DATA.—The calculation of the difference limen and of the discriminative sensitivity is similar to that in the preceding test, save, of course, that the standard is now 100, instead of 80 g.

RESULTS.—(1) *Normal capacity.* The discriminative sensitivity for cutaneous pressure depends so largely upon the type of the instrument (including especially the area of the pressure stimulus and the manner of application) that the norms obtained with other instruments can not be assumed to hold good for the present form of balance. Jastrow's results indicate a constant of approximately 1/15, which is nearly equal to that for lifted weights. Merkel similarly, reports 1/14 for his pressure balance, though Griffing believes that so fine a capacity as this must be attributed to the presence of a "muscular reaction of the finger." With a standard of 100 g. applied to the palm of the left hand, Miss Thompson found limens ranging from 4 to 20 g.

(2) *Dependence on the standard pressure.* The limen for the same *S*, with the same instrument and method, is constant, at least for stimuli between 50 and 2000 g. (Weber's Law).

(3) *Dependence on the area of stimulation.* According to Külpe (6, p. 160), the limen is 1/19 to 1/20 with an area of contact 1 mm. in diameter, but rises to from 1/13 to 1/16 with an area of contact 7 mm. in diameter.[3] Griffing, however, de-

---

[1]The author has found that some *E*'s, by a curious kind of unconscious 'sympathy,' are inclined to apply the lighter pressure more gently than the heavier pressure. *S*'s judgments will almost certainly, even without his knowledge of it, betray the operation of this secondary criterion by exhibiting an unexpected and impossible delicacy of discrimination.

[2]If it were not for the awkwardness of the position, it would, perhaps, be better to insert the finger volar side uppermost, in order more certainly to ensure against this movement of reaction in unreliable *S*'s.

[3]The tips of the author's balance are 8 mm. in diameter.

clares that "the area of stimulation does not, on the whole, affect the accuracy of discrimination for weights, but individual peculiarities appear in the results obtained."

(4) *"Practise* seems to aid discrimination at places not accustomed to pressure stimuli" (Griffing).

(5) There is no constant *sex* difference (Dehn, 2, and Thompson, 9).

(6) *Dependence on length of interval.* Accuracy of discrimination does not vary appreciably when the interval between application of the two stimuli is prolonged to 10 sec. (Griffing) or even to 30 sec. (Weber).

(7) *Dependence on place stimulated.* For weights of 100 g. or more, there is no appreciable difference in the discrimination of pressure on the palm of the hand, back of the hand, and the volar side of the index finger, though the last is probably more sensitive for very light weights (Griffing).

(8) *Constant error.* Most *S*'s show a tendency, frequently a marked tendency, to overestimate the second weight, *i.e.*, to judge it to be heavier (Griffing).

(9) *Direct judgments.* The impression of the standard stimulus not infrequently becomes so clear that it is carried over from one trial to another, so that, at least with large stimulus-differences, *S* may pass judgment when the first pressure is applied.

REFERENCES

(1) T. L. Bolton and Donna L. Withey, On the relation of muscle sense to pressure sense. *NebraskaUniv.Studies*, 7: No. 2, April, 1907, 175-195.

(2) W. Dehn, Vergleichende Prüfung über den Haut- und Geschmacksinn bei Männern u. Frauen verschiedener Stände. Dorpat, 1894.

(3) J. A. Gilbert, Researches upon school children and college students. *UnIowaSdPs*, 1: 1897, 1-39.

(4) H. Griffing, On sensations from pressure and impact. *PsMon*, 1: 1895, No. 1, pp. 88. (Also *ColumbiaConPhPsEd*, 4.) Summarized in *PsR*, 2: 1895, 125-130.

(5) J. Jastrow, On the pressure sense. *AmJPs*, 3: 1890, 54-6.

(6) O. Külpe, Outlines of Psychology. Eng. tr., London, 1895.

(7) J. Merkel, Die Abhängigkeit zwischen Reiz. u. Empfindung (II). *PhSd*, 5: 1889, 245.

(8) E. C. Sanford, A course in experimental psychology. Boston, 1895 and 1898. Pp. 449.

(9) Helen B. Thompson, The mental traits of sex. Chicago, 1903. Pp. 188.

(10) E. B. Titchener, Experimental psychology. Vol. II, Quantitative experiments. New York, 1905.

(11) G. M. Whipple, New instruments for testing discrimination of brightness and of pressure and sensitivity to pain. *JEdPs*, 1: 1910, 101-106.

TEST 22

Sensitivity to pain.—The determination of the threshold or limen for pain has been conducted for the usual comparative purposes, but it has had, in addition, a peculiar interest for some investigators, because it has been assumed that the limen varies in a characteristic manner with sociological status.

For the determination of the pain limen, use has been made both of electrical and of mechanical stimulation. Electrical stimulation (induction-coil current) has been employed chiefly by the Italian criminologists: pressure stimulation (upon the temple, palm, or finger-tip) has been employed almost exclusively by more recent investigators, and to this form of test our attention will be chiefly confined.

The value of the conclusions that have been so far reached from the use of pain tests is minimized by the difficulties, not always clearly realized, which appear in their administration. These difficulties, like those of most functional tests, arise primarily from the presence of a number of variable factors. The most important of these factors are: (1) dependence of the limen upon *S*'s ability to keep pain distinct both from strong pressure and from simple discomfort, (2) dependence upon the rate of application of the stimulus, including the length of time elapsing between successive applications, (3) dependence upon the place of stimulation, (4) dependence upon the area of the stimulus, (5) dependence upon the general condition and attitude of *S*, his "good-will," degree of fatigue, amount of practise, etc., (6) dependence upon individual constitutional differences in sensitivity, including sex, age, etc. This last is, of course, the particular dependence sought for in the results; the others, then, constitute disturbing factors and must, accordingly, be eliminated or at least evaluated. Proofs of these several dependences are given below in the discussion of results, but the first and second of them demand consideration here because they determine the choice of apparatus and of method.

(1) The dependence upon *S*'s judgment as to what constitutes pain has been recognized by most investigators as the primary source of difficulty in this test. It is undoubtedly true that pain is a specific sensory quality, distinct from pressure and distinct from unpleasantness; yet it appears equally true that it is often difficult, even for a practised *S*, to disentangle from his experience the three elements—cutaneous pressure, cutaneous pain, and discomfort. Children, to say the least, are not always competent to make such a differentiation, at least with the method of procedure that has commonly been followed.

To meet this difficulty, some *E*'s have instructed their *S*'s to wait for the distinct appearance of pain; others have asked them to report the first appearance of discomfort—an affective experience that might, or might not, be accompanied by pain, and which can thus scarcely be regarded as a rational index of the real pain limen. Thus MacDonald says (14, 15): "As soon as the subject feels the pressure to be in the *least disagreeable*, the amount of pressure is read from the scale . . . . The subject sometimes hesitates to say just when the pressure becomes the least disagreeable, but this is part of the experiment (!). The idea is to approximate as near as possible to the threshold of pain." Griffing, however, thinks that there is little liability to error from this source: "It is very easy to tell," he says, "when the pressure begins to be uncomfortable, and the 'imagination' does not seem to be a disturbing factor. Indeed, the pain seems often to come with greater [great?] suddenness." We may wonder, then, why he says in the preceding paragraph: "The observers were asked to speak when the instrument began to hurt at all or to be uncomfortable; for it was found that individuals differed as to what they called 'pain.'" Perusal of the literature makes it evident, as these instances illustrate, that some *E*'s have been measuring a "discomfort" limen, other a "pain" limen: doubtless, in either case, some *S*'s reported "discomfort," while others reported real "pain."

Yet again, it appears that schoolboys have sometimes understood the test to be a measure of their endurance of pain, and have manfully asserted: "It doesn't hurt yet" when the pain limen has long been exceeded.

This difficulty of identification of the pain consciousness cannot be wholly avoided, but it may be met, in part by giving *S* a clear account of the experience he is to report as pain, in part, especially in doubtful cases, by repeating the test, and in part by comparison of the results of a given *S* with the established norms for individuals of his age, sex, and type.

(2) Unless the rate of application of the pressure is constant from test to test, there is introduced a serious variable error. Roughly speaking, the limen will be higher if the pressure is applied rapidly, lower if it is applied slowly. In the use of the ordinary type of pain-tester, *e. g.*, of Cattell's (see 15, p. 1161) or of MacDonald's (13 and 15, pp. 1155-6) algometer, the rate of application of pressure is difficult to control; moreover, the rate has never been standardized, so that different investigators have followed different rates.[1] The chief merit of the pain-balance, or balance-algometer, as employed by Gilbert or as prescribed in the present test, consists in the guarantee that it affords of a rate of pressure increase that shall be uniform from step to step during each trial and from trial to trial.

Again, the time interval between successive trials must be standardized. If a given region, say the right temple, has been tested, its sensitivity is

[1] In illustration, Griffin applied pressure with the Cattell algometer at the rate of 1.4 kg. per sec., whereas Gilbert applied pressure with his own instrument at the rate of 50 g. per sec., or only 1-28 as fast.

increased for some time thereafter, yet some investigators have not hesitated to make a series of 5 or 6 tests upon the same spot in immediate succession. On the other hand, if a given region be subjected to daily tests for several weeks, its sensitivity becomes reduced by a process of inurement. It is clear that both of these sources of error must be avoided in the determination of the limen.

The remaining sources of error are fully illustratd below, and are intelligible without further discussion. When all precautions have been taken, it is probable, however, that the results of this test will be more variable and less reliable than those of other psychophysical tests.[1]

APPARATUS.—The author's (21) pressure-pain balance (Fig. 53). Cardboard screen with suitable supports. Seconds' pendulum (Fig. 28) or other device for time-control. A low table. [Telegraph sounder (Fig. 33), battery and wire.]

PRELIMINARIES.—Arrange the instrument and screen as in Test 21. Place the pendulum within easy range of vision, or, since the time-relations are so important, convert it into an auditory signal by the use of the telegraph sounder and battery (preferably adjusted to give a rather faint click).

METHOD.—Seat *S* comfortably so that his hand lies upon the hand-rest with his finger-tip between the pressure-tips of the instrument, as described in Test 21. Give him the following instructions: "I want to measure your sensitiveness to pain. There is nothing for you to be afraid of, as I will stop the moment you tell me that you notice any pain. I shall add these weights, one after the other, on the end of this bar, and I want you simply to notice what you feel in your finger-tip. The pressure will grow stronger, bit by bit. It will, perhaps, feel uncomfortable after a time, but never mind that. Wait for the first moment when it really hurts, when you feel a stinging, sore feeling, or a real ache. Do you understand what I mean? I don't want to know *how much* pain you can 'stand' without crying out; I don't want to know when it is simply *uncomfort-*

[1]The comments just given make it evident that MacDonald's algometer and Cattell's algometer are inadequate instruments. It follows that the results published by Griffing, Wissler, Swift, MacDonald, and Miss Carman are of doubtful value. Gilbert's results, though obtained by a better instrument, are more uneven than those of any other of the tests that he undertook and, on account of the slow rate of application that he used, are not directly comparable with those obtained by the method outlined below.

able; I want to know when you first notice what you would call *actual pain.*"

Throw the release-lever up to the left, so that the support beneath the balance-beam is permanently depressed: this makes possible a continuous, but cumulative pressure upon the finger.

Apply one of the large brass disc-weights, marked *B-200 g.,* every 2 sec. These discs are placed, *without jar,* upon the pin marked *B,* at the outer end of the beam. Continue application until *S* reports pain, then immediately remove the pressure. The total weight on the beam measures *S*'s pain limen.

Repeat the test with the left forefinger.

Variations of Method.—(1) Test other fingers of both hands.

(2) Apply the stimulus weights at a slower rate, say once in 4 sec., and note the effect upon the limen.

(3) Substitute a series of pressures for the cumulative, continuous pressure, by using the release-lever as in Test 21, and applying the pressure for 1 sec. only, after each weight is added.

Results.—(1) The short series of tests available with the pressure-pain balance thus far has shown that, with adults, the pain limen may be expected to lie between 1600 and 2400 g., when the method of cumulative pressure is employed.

(2) *Dependence on rate of application.* (*a*) "The rate at which pressure is added influences greatly the amount that is required to produce pain" (Gilbert).

(*b*) Immediate or close repetition of stimulation causes increased sensitivity, but continued practise for several weeks appears to reduce it (Griffing 9).

(3) *Dependence on nature of stimulus.* Sensitivity to pain produced by electrical stimulation bears no noticeable relation to that produced in the same person by pressure stimulation (Griffing, 8).

(4) *Dependence on the area of the stimulus.* "The pain threshold increases with the area of stimulation in an approximately logarithmic proportion" (Griffing, 9). Thus, when Cattell's algometer was applied to the palm of the hand, areas of 10, 30, 90 and 270 sq. mm. were correlated with limens of 1.4, 2.8, 4.4, and 6.6 kg., respectively.[1]

[1]The area of the pressure tip in the author's balance is approximately 50 sq. mm.

(5) *Dependence on region.* The regions of the body most sensitive to pain (from pressure stimulation) are those over the frontal and temporal bones, while the heel, the back, and the muscular regions of the leg and the hand are distinctly less sensitive. Illustrative limens obtained by Griffing are:

	KG.		KG.
Top of head	= 1.8	Right thigh, ventral surface	= 4.3
Forehead	= 1.3	Left hand, volar side	= 6.2
Right temple	= 1.0	Right heel, plantar side	= 7.0
Left temple	= 1.3	Back	= 8.0

These differences seem "to depend largely upon the thickness of the skin and the extent of the subcutaneous tissues." The left side of the body is in general somewhat more sensitive than the right. Differences of sensitivity will be found in the several fingers and frequently between corresponding fingers of the two hands.

(6) *Dependence on sex.* Gilbert (7), MacDonald, Dehn (5), Carman (3), Swift (17), and Wissler (22) agree that women are more sensitive to pain than are men. Thompson (18) agrees to this generalization, but adds that there are more men than women with very low thresholds, *i.e.*, that there is greater variability in men. Ottolenghi (16) and Lombroso (11), on the other hand, state that (with electrical stimulation) women are markedly less sensitive than men.

The latter authority believes that this result is confirmed by the experience of surgeons, who find that women possess greater endurance of pain: the popular opinion that women are more sensitive to pain is due, in his view, to the greater tendency of women to express feelings of pain by tears or otherwise; he also believes that their greater longevity may be due partly to their inferior susceptibility to pain.[1]

Typical results are those of Wissler and of Gilbert: the former publishes the following averages (Cattell algometer on the ball of the right thumb); college men, 5.9 kg.; college women, 2.4 kg. Gilbert finds that the average difference between boys and girls is about 400 g., and that this difference increases with age, until at 18 or 19 it becomes over 1 kg. (See Table 30.)

[1]Cf. Burt and Moore, The mental differences between the sexes. *JEPd*, 1: 1912, p. 366.

(7) *Dependence on age.* Sensitivity to pain, in general, decreases with age up to 18 or 19 years, and is thenceforth approximately stationary, but Carman and MacDonald both find irregularities near the period of puberty, and Wissler finds Seniors more sensitive than Freshmen (as a class). It is probable that the general result is disturbed by a tendency on the part of younger children to shrink from the test and to report discomfort rather than pain. Gilbert's results are embodied in Table 30.[1].

TABLE 30.

Pain Limen, in kg., for about 50 Boys and 50 Girls of each Age (Gilbert).

AGE	6	7	8	9	10	11	12	13	14	15	16	17	18
Boys	1.26	1.38	1.70	1.69	1.67	2.07	2.00	2.05	2.13	2.35	2.70	2.75	2.85
Girls	1.15	0.93	1.18	1.36	1.45	1.56	1.46	1.70	1.82	1.77	1.85	1.93	1.80
Average	1.21	1.16	1.44	1.53	1.56	1.82	1.73	1.88	1.98	2.06	2.28	2.34	2.33

(8) The range of *individual difference* is large in the pain test. Gilbert, for example, found that the mean variation for his groups of children ranged from 330 to 820 g.

(9) *Dependence on fatigue.* Swift (17) concludes that fatigue increases sensitivity to pain, especially in the case of younger pupils and of girls, because it lowers the tone and increases the irritability of the whole system.[2] Essentially similar results are reported by Vannod (19) and by Vaschide (20). The former employed an instrument ('algesiometer') analogous to v. Frey's 'hair-esthesiometer,' and found that the pressure needed to produce pain fell from 45 g. at 8 a.m. to 39 g. at 10 a.m., and to 29 g. at 12 m., under the influence of school work. The latter concluded that pain tests warranted him in stating

[1]It should be remembered, again, that Gilbert used a very slow rate of application, so that his results may not be comparable with those obtained by the methods we have prescribed.

[2]This conclusion is based upon tests before and after a 10 day's vacation in which the "physical condition" was determined by a dynamometer—a method already shown to be of doubtful value (Tests 6 and 9). There is no evidence to indicate that check-tests were made to determine the range of variations that might have appeared under constant work-conditions.

(1) that mathematics and ancient languages possess an especially high fatigue-value, (2) that written exercises in the form of tests produce intense intellectual fatigue, (3) that afternoon is much more fatiguing than forenoon instruction, and (4) that a forenoon spent outside of the school permits a return, in most cases, to normal sensitivity. On the other hand, Binet (1), who used an adaptation of Blocq's sphygmometer, has come to diametrically opposite conclusions, and asserts that the effect of fatigue is to reduce, not to heighten, pain sensitivity.

(10) *Dependence on mental ability.* Binet and Simon (2, p. 58) declare that the more intelligent children have a lower pain limen. Carman also found that bright boys (teacher's estimate) were more sensitive than dull boys. Swift comes to a similar conclusion by contrasting the best with the poorest fifth of a class, and attributes the result to the more delicate nervous organization of bright children. MacDonald, however, says "there is no necessary relation between intellectual development and pain sensitiveness." "Obtuseness to pain seems to be due more to hardihood in early life."

A curious and somewhat dubious correlation unearthed by Miss Carman is that boys and girls who are especially dull in mathematics are more sensitive on the right than on the left temple.

(11) *The feeble-minded.* Scientific tests of the sensitivity of feeble-minded children present obvious difficulty: it is hard, as we have seen, to induce even normal children to report accurately the first appearance of the pain quality. However, everyday observation, as well as cruder tests of reaction to pinching, burning, etc. (see especially Binet and Simon), would indicate that the feeble-minded, as a class, are markedly insensitive to pain.

(12) *Dependence on sociological condition.* (*a*) Somewhat unconvincing is the series of conclusions in which MacDonald (12, 14, 15) summarizes his correlations between pain sensitivity and sociological condition; *e.g.*, girls in private schools, who are generally of wealthy parents, are more sensitive than girls in public schools; university women are more sensitive

than washerwomen, but less sensitive than business women; self-educated women are more sensitive than business or university women (owing, perhaps, to overtaxing their nervous systems in the unequal struggle for an education); the nonlaboring classes are more sensitive than the laboring classes, etc.[1]

(*b*) The study of the pain sensitivity of the *criminal* is a specific sociological problem that has attracted much attention since the concept of the 'criminal type,' or of the 'instinctive criminal,' was introduced by Lombroso and his school. It has been generally stated that the typical criminal is distinctly less sensitive to pain than the average normal man, and it has frequently been added that the moral insensibility of the criminal is to be largely attributed to this bodily insensibility. These statements are based upon certain experimental tests and upon common observations of the hardihood and general obtusity of the 'typical criminal.' Nevertheless, recent pain measurements indicate that the generalization is too sweeping, and that numerous exceptions occur. It may even be doubted whether the existence of a distinct criminal type has been satisfactorily established.

A general summary of the pain sensitivity of criminals is given by Ellis (6, Section 8). The inadequacy of the algometer test as applied to criminals is discussed briefly by Miss Kellor. A typical exception to the general belief is found in Dawson's conclusion (4) that normal children are less sensitive to pain than delinquent children, probably because many of the delinquents were of neurotic type.

(13) *Miscellaneous correlations* reported by Miss Carman are: boys with light hair and eyes are less sensitive than boys with dark hair and eyes. First-born are more sensitive than second-born boys, and the latter than later-born brothers: the same is true of girls, save on the right temple (!). These conclusions are subject to obvious criticism.

[1] The measurements from which these conclusions are drawn were made by different investigators, by an unreliable method, and have been assembled apparently by mere comparison of averages and with no attempt to determine the limit of error; they might, or might not, be confirmed by more exact methods.

REFERENCES

(1) A. Binet, Recherches sur la fatigue intellectuelle scolaire et la mesure qui peut en être faite au moyen de l'esthésiomètre. *AnPs*, 11: 1905, 1-37, especially 32 ff.

(2) A. Binet and T. Simon, L'intelligence des imbéciles. *AnPs*, 15: 1909, 1-147 (section on Le sens de la douleur, 52-8).

(3) Ada Carman, Pain and strength measurements of 1507 school children in Saginaw, Michigan. *AmJPs*, 10: 1899, 392-8.

(4) G. E. Dawson, A study in youthful degeneracy. *PdSe*, 4: 1896, 221-258.

(5) W. Dehn, Vergleichende Prüfung über den Haut- und Geschmacksinn bei Männern u. Frauen verschiedener Stände. Dorpat, 1894.

(6) H. Ellis, The criminal. 3d ed., London, 1907.

(7) J. A. Gilbert, Researches upon school children and college students. *UnIowaSdPs*, 1: 1897, 1-39.

(8) H. Griffing, On individual sensitivity to pain. *PsR*, 3: 1896, 412-5.

(9) H. Griffing, On sensations from pressure and impact. *PsMon*, 1: 1895, No. 1. Pp. 88. (Also *ColumbiaConPhPsEd*, 4.) Summarized in *PsR*, 2: 1895, 125-130.

(10) Frances Kellor, Experimental sociology. N. Y., 1901. Pp. 316.

(11) C. Lombroso, The sensibility of women, brief report, *Mind* n. s. 1: 1892, 582.

(12) A. MacDonald, Sensibility to pain by pressure in the hands of individuals of different classes, sexes and nationalities. *AmJPs*, 6: 1895, 621-2.

(13) A. MacDonald, A temporal algometer. *PsR*, 5: 1898. 408-9.

(14) A. MacDonald, Further measurements of pain. *PsR*, 6: 1899, 168-9.

(15) A. MacDonald, Experimental study of children, etc., reprint chs. 21 and 25, *RepComEd*, 1897-8, Washington, 1899.

(16) S. Ottolenghi, La sensibilité de la femme. *RSci*, Ser. 4, vols. 5: 395, and 6: 698.

(17) E. Swift, Sensibility to pain. *AmJPs*, 11: 1900, 312-7.

(18) Helen B. Thompson, The mental traits of sex. Chicago, 1903. Pp. 188.

(19) Th. Vannod, La fatigue intellectuelle et son influence sur la sensibilité cutanée. *RMdSuisse*, 27: 1896, 21.

(20) N. Vaschide, Les recherches expérimentelles sur la fatigue intellectuelle. *RPh*, 5: 1905, 428.

(21) G. M. Whipple, New instruments for testing discrimination of brightness and of pressure and sensitivity to pain. *JEdPs*, 1: 1910, 101-106.

(22) C. Wissler, The correlation of mental and physical tests. *PsMon*, 3: No. 6, 1901. Pp. 62.

TEST 23

Discrimination of dual cutaneous impressions: Esthesiometric index.—As long ago as 1834, E. H. Weber, a German physiologist, observed (62) that, if two punctiform pressures are applied simultaneously to adjacent points on the skin, a single impression results, whereas, if the pressure points are applied

at gradually increased distances, an extent can be discovered which is just sufficient to yield a perception of two points. Weber explored many regions of the skin and published extended tables of measurements of this distance, which has since become known variously as the "limen for duality," or "doubleness," as the "esthesiometric index," the "space threshold," or even, less exactly, as the "index of delicacy of touch."

On account of Weber's explanation of the phenomenon, which was in terms of the supposedly quasi-circular distribution of the end-organs of the sensory nerves, the experiment is often referred to as the test of "sensory circles." On account of the type of instrument employed, it is sometimes termed the "compass test."

Since Weber's time the experiment has become a classic in psychology. Seemingly simple and definite, more careful examination has revealed the fact that the determination of the esthesiometric index is in reality unusually difficult, and that the factors which underlie the observer's judgment are surprisingly varied and subtle.

For differential psychology, the chief interest in the test is found in its use by criminologists to measure "general sensibility," and by several German investigators to measure the degree of fatigue of school children. Physicians, also, have employed it for diagnostic purposes, particularly in connection with pathological conditions of the spinal cord, and it has found special favor in the psychological laboratory, both for its intrinsic interest and for the illustration of various psychophysical methods.

As in the case of other tests, the chief difficulty in the use of the esthesiometric test lies in the presence of numerous sources of error, which must be fully recognized and controlled if valid results are to be secured. In general, it may be said that the esthesiometric limen will depend upon (1) the instrument employed, (2) the region of the body tested, (3) the method of procedure, including the nature of the instructions, (4) the care with which *E* applies the stimulus and the actual pressure employed, (5) *S*'s degree of fatigue, (6) *S*'s degree of practise, (7) *S*'s ability to attend to the impressions and to make

accurate reports, and his general type of judgment, especially in the 'critical region,' and (3) upon a number of other factors, such as *S*'s sex, age, the condition of the circulation in the region tested, etc. The manner in which these factors affect the index will be discussed below.

The instrument employed may be extremely simple, *e.g.*, the set of needles thrust through bits of cardboard, used by Binet in his earlier tests (2) and subsequently improved by Buzenet (14) and by Hill (26), or it may be very complicated and elaborate.

In general, the development of the esthesiometer since Weber's time has been in the direction of greater complexity and delicacy, with a view of affording more adequate control of the separation of the points, of the simultaneity of their application, and of the degree of pressure exerted. It is doubtful whether much of this elaboration is needful: objective equalization of the pressure does not insure subjective equalization, and a careful *E* is better able to apply the points simultaneously if he works with a relatively simple instrument.

The instruments selected, an improved form of Jastrow's esthesiometer for one method and Hill's needles for another, possess all the requisite features. For other models, consult Blazek (7), Binet (3, 4), and Washburn (60). The models of Ebbinghaus and v. Frey are figured in Zimmermann's catalog. Spearman's instrument is described in Sommer (50) and pictured in use in Schulze (44, p. 67).

Two methods are described, each with the apparatus best adapted to it: the first, or ordinary method follows more closely the regular procedure of the laboratory and is to be considered standard for competent *E*'s and adult or well-trained *S*'s. The second is better adapted for less skillful *E*'s and for use with children or untrained adults.

A. ORDINARY METHOD

APPARATUS.—Jastrow's improved esthesiometer (Fig. 55). Cardboard screen and supports. Pillow or folded towel.

METHOD.—(*a*) *Preliminary practise.* Seat *S* comfortably with his right forearm laid horizontally, volar side uppermost, upon a small pillow or folded towel, with the clothing arranged to expose the forearm from elbow to wrist, without impeding the circulation at the elbow.

With ink or pencil mark with a transverse line the region approximately midway between the wrist and the elbow.

Arrange the screen to cut off from *S* the view of his forearm and of the instrument.

Devote from 2 to 5 min. to a preliminary practise series in order to familiarize *S* with the test, particularly with the perception of one and of two points.

Instruct all *S*'s, in the same words, substantially as follows: "I'm going to touch your arm with points, something like pencil-points. They won't hurt you at all. You are to give careful attention to what you feel, and tell me if you think I'm touching you with one point or with two points. You will have to watch very carefully. If you feel only one point, say 'one;' if you feel two, say 'two.'"

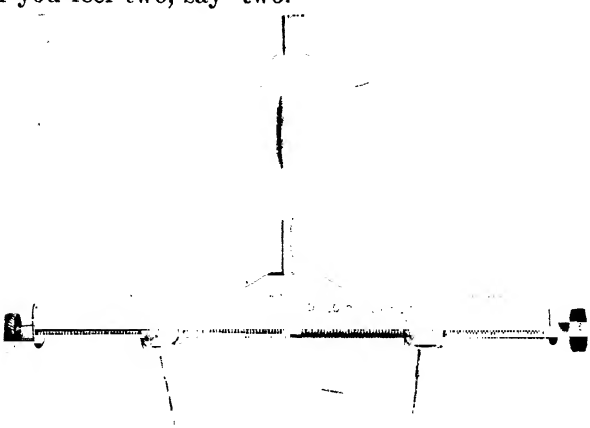


FIG. 55. JASTROW'S IMPROVED ESTHESIOMETER.

Begin with a distinctly supraliminal distance, say 80 mm. After a warning 'ready' signal, bring the instrument down perpendicularly upon the middle of the forearm, parallel with its longitudinal axis, in such a manner that both points make contact simultaneously,[1] and rest by their own weight upon the skin, as the holder is allowed to slide one or two cm. down the stem of the instrument. The several applications should be uniform in duration—about 1.5 sec.

[1]This is imperative because the limen for successive stimuli is only 1/8 to 1/4 that for simultaneous stimuli.

Next apply two points at 70 mm., then 60 mm., and so on, with occasional trials of one point. Leave an interval of 7 to 10 sec. between trials to allow the preceding sensory disturbance to die away. Keep the applications in the same general region with one point of the compasses below and the other above the transverse line, but do not seek to apply the compass upon exactly the same spots at each trial. When one point is applied, place it in the neighborhood of the spots where one or the other of the two points are being applied. Avoid contact with hairs, which set up tickling sensations, or pressure upon projecting veins or tendons, which will be of a character dissimilar to the normal contact.

This practise should help *S* to be familiar with, and to distinguish the 'feels' of one point and two points. It will be found that some *S*'s will, nevertheless, occasionally answer "two" when but one point is given. This is the not uncommon *Vexirfehler*, or esthesiometric paradox, which is well recognized as a source of difficulty in esthesiometry. If it proves persistent, it will probably be impossible to determine an exact limen with that particular *S* unless Method B proves more successful. To avoid it, *E* may allow *S* to look at the instrument when one point is resting on the skin and he has just announced "two." Sometimes young *S*'s, who have caught sight of the instrument, may, with childish logic, conclude that there must always be two points because there are two on the instrument. Some investigators never permit *S* to catch sight of the instrument on this account.

Children who display timidity must be encouraged to adopt a more favorable attitude toward the test.

(*b*) *Test proper.* Allow *S* a short rest (during which his arm should be withdrawn from the somewhat constrained position); then proceed from a distinctly supraliminal distance toward the critical region until *S* judges "one," and follow this with an ascending series from clearly "one" until *S* judges "two." Average the two determinations for the limen.

B. METHOD OF CONTRAST

As is explained more fully in the results that follow, the esthesiometric experiment is often rendered difficult, first by the circumstance that a real illusion of two points may arise when one point is applied, and secondly, by the circumstance that there really is no abrupt transition between the clear perception of two points and the clear perception of one point. Rather there exists an intermediate region where neither

clearly one nor clearly two is felt. Evidently, these intermediate experiences may be reported by some *S*'s as "one," by others as "two;" and indeed, the same *S* may shift his criterion of response during the course of an experiment. Various plans have been proposed for circumventing this difficulty. It is possible, for instance, to instruct *S* to answer "one" unless he feels two quite clearly distinct points (Rivers, Burt). Another plan, which has been followed with some measure of success by

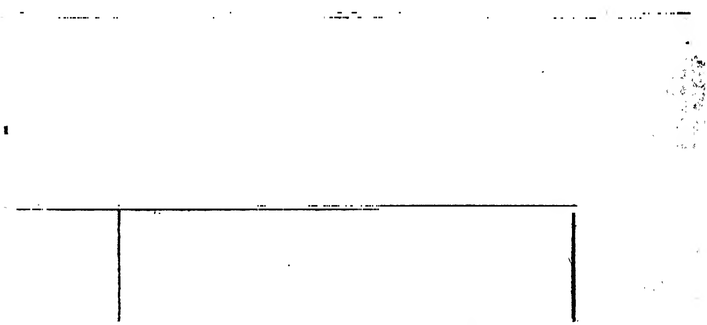


FIG. 56. NEEDLE ESTHESIOMETER.

Wagner, Binet, Martin and others, consists in applying the stimuli in such an order as to produce as much contrast as possible between the feel of one point and the feel of two points. If, in addition, the one-point stimuli and the two-point stimuli are applied with equal frequency, any tendency of *S* to favor the one or the other judgment is much reduced.

To make the rapid changes that are necessitated by this method in the distances separating the compass-points it is better to use special apparatus, *i.e.*, apparatus that will provide

a series of different fixed distances, as has been recommended by Binet and Henri, Meumann, Buzenet, Hill and others.

APPARATUS.—Set of 14 needle esthesiometers (Fig. 56), after Binet and Buzenet, modified by Hill. These afford 13 separations from 10 to 70 mm. by 5 mm. steps and one single point (zero separation).

METHOD.—Preliminary arrangement of *S*'s arm, screen, etc., and instructions to *S*, as in Method A.

Show *S* the apparatus, so that he knows that either one point or two points may be applied. Hold the esthesiometers by the metal handle; apply them gently and evenly to the skin and let them rest by their own weight, merely steadying them by the handle. Work downward from the widest separation toward the threshold, using the instruments always in pairs containing one double point and the single point, *e.g.*, 70 and 0, 65 and 0, 60 and 0, etc., in other words, so that one point is applied as often as two points, and so that clearly one and clearly two are frequently perceived in close succession. Apply each contrasting pair two or three times in irregular order, and reduce the separation of the double-point stimulus until a separation is reached at which *S* calls the two points "one."

This separation is now applied, together with the zero separation, in a series of 20 trials (10 zeroes and 10 doubles arranged in chance order). If *S* makes more than two errors with the 10 double points, try a new set of 20 applications with the next larger separation of the two points; if but one error or no errors, a set with the next smaller separation of the two points. The threshold may be taken as the distance at which two errors are first made with the ten double points, unless subsequent better records with lesser separations show that these errors were due to a temporary lapse of attention.

VARIATIONS OF METHOD.—Test other regions of the body. Use a transverse application on the forearm. Compare the sensitivity before and after periods of rest or of fatiguing work. Test the acquisition of practise and its transfer to symmetrical and adjacent portions of the body. Try the method advocated by Rivers for uncultured *S*'s of informing *S* after each response whether he is right or wrong.

Results and Conclusions.[1]—(1) *Dependence on sex.* Wissler (63) could discover no sex differences in cutaneous discrimination of dual pressure, but Miss Thompson (53), despite some acknowledged difficulties in the administration of her test, concludes that "women have a somewhat finer discrimination in the crosswise direction, and a decidedly finer discrimination in the lengthwise direction." Actual figures cited are 20 and 65 mm. for women and 35 and 75 mm. for men, in the transverse and longitudinal directions, respectively.

Burt and Moore (13) figure the differences found by Miss Thompson as showing an inferiority of the men of 32 per cent. in lengthwise and 7 per cent. in the transverse direction (that is, 32 per cent. of the women surpass the median of men, etc.). Burt further cites the work of Galton, who tested over 1200 persons on the nape of the neck and found for the men an average of 14.04 (range 6 to 19) and for the women 13.70 (range 4 to 15). Again, Burt and Moore declare that the sex difference in esthesiometry is the largest they have discovered in any mental test, being so large (—47.7 per cent. in one group of school children and —37.4 per cent. in another, in terms of proportion of males exceeding the median of females) that the curves for the two sexes scarcely overlap.[2] They regard this difference, therefore, as peculiarly striking and undoubtedly innate. Lombroso seems to be the only writer of importance whose results disagree with this conclusion, and his technique was so primitive as to throw decided doubt upon the validity of his results.

(2) *Dependence on age.* There is fair agreement, especially among the earlier investigators (see *e.g.*, Czermak, 15) that children have a greater sensitivity than adults. Similarly, Wissler found Seniors inferior to themselves as Freshmen. In regions where the sensitivity is poor, the difference between

[1]In reporting these results, it may be stated that many of them have been obtained by methods that are open to criticism, particularly in that suitable precautions have been wanting to control or to eliminate the numerous disturbing factors already mentioned.

[2]As a matter of fact, a larger sex difference was found for lifted weights. See Test 20, Result 4.

children and adults is quite distinct, *e.g.*, a limen of 67 mm. on the thigh of an adult in contrast to 35 mm. upon the same region in a boy of 12.

The usual explanation that the child has a greater number of nerve endings within the same-sized area is probably only a partial explanation, as it has been computed that the child's sensitivity is out of all proportion to the differences in dimensions here concerned. A contributory factor is doubtless to be found in the fact that the child's skin is thinner and more tender, so that a given impact produces a sharper sensory experience.

Griesbach (20), who found no difference between children from 11 years up to 19, is practically the only investigator who has not found the child more sensitive than the adult, but Griesbach's figures can not be accepted without misgivings.

(3) *Dependence on region.* (*a*) Table 31, which is derived from Weber's original results, gives an idea of the topographic distribution of sensitivity.

TABLE 31.

Topography of Esthesiometric Sensitivity (*Weber*)

REGION	LIMEN IN MM.	REGION	LIMEN IN MM.
Tip of the tongue	1.1	Forehead	22.5
Tip of the fingers	2.2	Back of hand	31.5
Mucous membrane of the lips	4.5	Forearm	40.5
End of the nose	6.7	Back	54.1
Cheek	11.2	Thigh	67.6

An empirical generalization, known as Vierordt's Law (57, p. 298), summarizes these differences in sensitivity, especially of the limbs and head, by the statement that the delicacy of discrimination of two regions on the skin of a portion of the body that is moved as a whole is proportional to the average distance of these regions from their common axis of rotation.[1] In illustration, if the discriminative sensitivity of the tip of the shoulder (acromion) be taken at 100, then that of the upper

[1] Somewhat analogously, Krohn (32) cites an instance in which a man whose arm had been held immovable in a plaster cast for three months exhibited marked decrease of sensitivity of that arm shortly after the cast had been removed.

arm is 151, of the forearm 272, hand 659, thumb 2417, middle finger 2582.

Burt's results (11) for 13-year-old school boys on the volar side of the forearm were for an Oxford elementary school, average 36.2 mm., M.V. 9.0, range 19-58.3, and for a high-class preparatory school, average 38.9 mm., M.V. 11.0, range 12.5-63.7.

(*b*) Van Biervliet (54), who asserts that in many sense-departments the favored *side of the body* is superior to the other side by the fraction 1/8, publishes tables which indicate that the same constant applies in esthesiometry.

(4) *Dependence on time of day.* A comparison of the sensitivity at different periods of the day, in the search for a diurnal rhythm, is, of course, complicated by the presence of fatigue (see the following paragraph), and doubtless by individual differences as well. Schmey (43) believes sensitivity to be less at night than in the morning. Adsersen (1), however, asserts that sensitivity is lowest in the morning, begins to improve at about 11, reaches a maximum from 3 to 7, and thence decreases. By comparison, he shows that this rhythm coincides very closely with the diurnal curve of bodily temperature, and he therefore argues that the esthesiometric fluctuations are indices of physiological changes common to the diurnal rhythm. Tawney's work (51), on the other hand, leads one to believe that the limen undergoes, in many *S*'s, such irregular fluctuations that it is impossible to find a limen that is constant for half an hour at a time.

(5) *Dependence on fatigue.* The effect of fatigue upon the limen forms the chief source of interest in connection with the esthesiometric test. Griesbach (20) was the first to make extended use of the test in the examination of school children. His amazingly uniform and definite results, when taken at their face value, indicate unequivocally that the method is of value and importance in the detection of the fatigue induced by school work. They have, moreover, been confirmed more or less thoroughly by Binet (6b), Blažek (7), Bonoff (10), Heller, Schuyten (46), Vannod (56), and Wagner (59). On the other hand, they have been controverted with equal emphasis by Bolton (8, 9), Germann (18), Kraepelin (31), Leuba (34), and Ritter

(40), while recent work by Miss Martin has given negative results. Investigations of fatigue by means of the esthesiometric test have also been made of late by Ferrai in Italy, Sakaki (42) in Japan, Ley in Belgium, Michotte in Belgium, and Noikow (38) in Bulgaria.[1]

Griesbach argues that fatigue reduces the power of sustained attention and that this in turn reduces the cutaneous sensitivity. He appears to believe that he is the first to have discovered this relation, but it may be noted that Weber, himself, had cautioned his readers to avoid fatigue if valid results were to be secured, that Schmey, in 1884, had demonstrated that the fatigue of the arm by calisthenic exercises reduced sensitivity, and that Stanley Hall (23), in 1879, had commented on the variability of the results obtained from Laura Bridgman, and had expressed the hope that "a curve of fatigue may be obtained by which some approximate comparison with the fatigue of a nerve-muscle preparation may be made." Griesbach worked on the glabella, cheek-bone, tip of the nose, under lip, ball of the thumb, and tip of the index-finger, and tested pupils before and after various kinds of school work, on Sundays, holidays, at the end of vacations, etc. A single example will suffice: a girl of 14 had a limen on the glabella of 5 mm. at 7 a. m., but this increased to 12.5 mm. at 12, noon, after a morning at school, whereas on Sunday her limen was but 3.5 mm. Griesbach concludes his memoir with a strong plea against overwork, and asserts that no schoolboy can meet in full the demands of present-day higher education without endangering his health (20, p. 88).

So far from being silenced by the critics of his method, Griesbach has subsequently published (1910) results which, he claims, demonstrate that esthesiometrical measurements of fatigue are adequate to explain the functions and localization of brain centers employed in various activities, and adduces such conclusions as: fatigue induced by mental or bodily activity does not affect both hemispheres alike. In mental work, especially linguistic and algebraical, different values afforded by the esthesiometer show that the left hemisphere is predominantly engaged in right-handed, the right in left-handed persons, while in bodily exertion, the right hemisphere is predominantly engaged in both right and left-handed persons, etc. So far as we are aware, no other investigator has confirmed these findings, nor has anyone even attempted to use the esthesiometric method for such precise inferential work.

Blazek divides pupils into three types: (1) those who possess ability, who work industriously and attentively, and thus exhibit distinct and progressive fatigue curves, (2) those who work intermittently, and whose curve is therefore broken by recuperative periods, (3) those whose curve is approximately a straight line. The fatigue curve, therefore, depends in the main upon the type of worker, but the individuality of the teacher and the subject-matter in hand are also determining influences. He concludes that more than half the pupils work irregularly and thus save themselves in part from overwork. He would recommend 4 subjects daily, of 45 min. each, with 15 min. rest periods between each subject.

Wagner says that the Griesbach method is a valuable adjunct for the study of fatigue. Afternoon instruction is practically valueless peda-

[1]For further discussion of the value of the method, consult Gineff (19), Meumann (37, vol. ii., 89-94, 107-110) and the recent monograph of Offner (39, pp. 31-38). Griesbach's rejoinder to his critics (21) should also be consulted.

gogically. Play and gymnastics are sources of fatigue to many pupils, and should be relegated to the close of instruction or to the afternoon. If the fatigue-value of mathematics be placed at 100, other subjects may be rated thus: Latin 91, gymnastics 90, geography and history 85, French and German 82, nature-study 80, drawing and religion 77.

Binet summarizes the results obtained by himself, and by a group of teachers who worked under his direction, by declaring that "intellectual fatigue is manifested by a reduction of sensitivity, measurable on the back of the hand: this reduction is revealed by fewer judgments of 'two' for smaller distances (0.5 to 1.5 cm.), is more pronounced for girls than for boys, and is to be attributed to an actual reduction of tactual sensitivity itself, not to a mere relaxation of attention" (6 b, p. 29). It is to be noted, however, that this conclusion is based upon the 'lump' results obtained from groups of school children: analysis shows that, in some groups at least, less than half of the pupils (*e. g.*, 31 of 75), gave evidence of fatigue in this way. In this investigation Binet employed an ingenious, though perhaps debatable method for measuring separately skin sensitivity and attention. The percentage of the answers "two" for separations 5, 10 and 15 mm. was used to indicate delicacy of touch, the percentage of the answers "one" for separations 20, 25 and 30 mm. to indicate relaxation of attention. His data show, after fatiguing school work, a reduction in the first, but no change in the second group of answers, and hence he argues that the fatigue of the school work reduced the sensitivity of the skin, but did not relax attention. It should be added that Binet claims no more than to have demonstrated the feasibility of the method for group tests.

Schuyten at first condemned the esthesiometric method (45), but later (46) he found to his surprise that it worked satisfactorily. He used groups of 5 selected pupils.

Ritter tested himself at intervals for two years, but could get no evidence of fatigue by the esthesiometer.

Bolton says the limen is so hard to determine that it cannot be satisfactorily accomplished in a single sitting. In his tests, severe mental work of two hours duration did not produce a measurable change in the index.

Germann, after making 2450 trials on a single *S*, could not discover any relation between fatigue and the limen.

Leuba admits that fatigue affects the limen, but says that it is only one of a great many factors. He was unable to draw any general inductions concerning fatigue, even when the results for three days of severe mental work were contrasted with those for three days of rest.

Meumann (37, vol. 2, 90 ff), similarly, admits that the tendency of fatigue is to heighten the limen, but protests vigorously that the relation is but indirect, and so complicated by numerous little-known factors that the numerical expression of shifts in the limen can in no wise be regarded as a measure of fatigue. ("Wir haben in der Erhöhung der Raumschwelle durch die Ermüdung *nur ein objectives Symptom derselben, aber keine Messung.*")

Kraepelin, however, declares flatly that investigations that embody measurements of fatigue by the use of the esthesiometer are "all in the air" (*stehen einfach in der Luft*), and are nothing but the unintentional expression of the preconceived opinions of the investigators.

(6) *Dependence on practise.* The effect of practise was studied by Czermak (15), as early as 1855, in his investiga-

tion of the sensitivity of the blind, was measured more carefully by Volkmann (58), and confirmed later in particular by Dresslar (16). From these investigations, it would appear that the practise-effect is visible within two hours, and may be pushed to unexpected lengths by continued work; thus, Dresslar reports one *S*, who started with a limen of 29 mm., reduced this to 21 mm. in the first week, to 10 mm. the second week, 5.5 mm. the third week, and 2.8 mm. the fourth week—a net reduction, then, to approximately 1/10 of the original figure. This practise-effect is, however, rapidly lost, being reduced very definitely within 8 days and completely lost within a month. The practise-effect is said to appear much more rapidly on fingers, hands, and other exposed parts, than on the back and other relatively inaccessible and immobile regions. Both Volkmann and Dresslar submit evidence to show that this 'education' is subject to *transfer to symmetrical regions,* though not to regions adjacent to the one on which it was effected.

In seeking an explanation for the effect of practise, we are met with the fact that other investigators have not confirmed the results just cited. Camerer, in his lengthy series, did not find such extensive practise-effects. Foucault (17) hesitates to commit himself on the effect of practise on the actual limen. Tawney (52) found his work so vitiated by *Vexirfehler* and by auto-suggestion in general that he was unable, for some *S*'s, to establish any constant limen. Both Tawney and Henri (24, b) deny that the influence of practise, when established, is confined to symmetrical regions of the body. Solomons (49) asserts that the practise-effect is rapid if *S* is informed of his errors, but practically non-existent if he is not—a principle which, if confirmed, may explain the disagreements just cited.

The explanation of the process of 'education' in this test is found by some writers to lie in certain peripheral processes—not necessarily in anatomical, since the practise is too rapid to admit of that, but rather in local physiological processes. Others, *e. g.*, Judd (28) and Solomons, believe that the education is essentially a 'central,' or psychological, process—an improvement in judgment due to the learning of new associations.

The low *threshold of the blind*, reported by Goltz, Gärtner, Heller, Miss Washburn (60), Hall (23), Jastrow (27), and others, is to be deemed a special example of practise, and does not imply the presence of exceptional sensitivity or special peripheral delicacy. Helen Keller (27) has a limen of 1.5 mm. on the tip of the left forefinger, and 3-4 mm. on the palm of the hand, which is smaller than that of the average *S*. Laura Bridgman (23) is credited with a limen of 0.7 mm. on the right forefinger, and her general sensitivity of touch, according to Hall, was "from two to three times as great as that of an ordinary person."

Foucault appears to be the only investigator who has not found the blind more sensitive.

(7) *Dependence on S's type of judgment.* Mention has already been made of the difficulty arising in esthesiometry owing to the different ways in which the perceptions intermediary between "one" and "two" may be reported.[1] Binet, for example, classified *S*'s as *simplistes, distraits* or *interpreteurs,* according to their mode of reporting these perceptions. But the most recent and valued critique of this field, that of Foucault, goes yet farther, and distinguishes eight possible perceptions and four types of *S*.

The eight different possible perceptions are:

(1) a single distinct point, without area,

(2) a clean-cut small circular area, with often a center where the pressure is stronger,

(3) like the 2d, but larger, though still circular,

(4) an elongated area of pressure, like a line, an oval, an ellipse, etc.,

(5) like the 4th, but with two points within it where the pressure is greater than in the region between them,

(6) two circular areas overlapping one another,

(7) two circular areas tangent to one another,

(8) two circles of pressure separated by a between-lying area of no pressure.

According to the manner in which they 'assess' or report these eight possible perceptions, there are, according to Foucault, four different types of *S*'s:

(1) the 'rational' *S*, who reports only 7 and 8 (of the perceptions) as "two," 1 and 2 as "one," and the others as "intermediate,"

(2) the 'prudent' *S*, who reports 1 to 6 as "one" and 7-8 as "two,"

(3) the 'bold' (*hardi*) *S*, who reports 1 to 4 as "one," and 5 to 8 as "two,"

(4) the 'imprudent' *S*, who reports anything save 1 and 2 as "two."

These types are found in actual work, and they may shift as the experiment progresses.

Foucault's way out of this complexity of issues [Binet is evidently quite right in saying that "esthesiometry ceases to be a simple little exercise"] is to train all *S*'s into the first, or rational type by giving them frequently, with knowledge, the feel of one point and two points and by coaching them to report "intermediate" whenever they experience Perceptions 3 to 6. He assures us that even young children can submit to this training, so that it does become possible, after all, to measure the threshold. In substance, it will be seen that his plan virtually means the

[1] The fact is, of course, that all the conditions of the experiment, and particularly the instructions and the sight of the apparatus conspire to force upon *S* what the psychologist calls the "stimulus error." *S* is constrained to answer either "one point" or "two points." Ideally, he should, perhaps, never see the apparatus, and be instructed to report accurately and in detail just precisely what sort of cutaneous experience he senses.

instruction of *S* to answer "two" when the two areas of contact do not seem to overlap one another. The frequent giving of two points means, of course, of two supraliminally distant points. Only series of stimuli in which such distances frequently occur can be called 'normal' or standard series, in Foucault's judgment. All other series are abnormal or misleading (*trompeuse*), and tend, in particular, to develop the paradoxical error—a statement that coincides well with the results of Solomons and of Tawney, both in the latter's own investigation (52) and in that undertaken with Henri (25). These writers have laid emphasis upon the role played by suggestion in esthesiometry. They were able, for example, by arranging conditions suitably, to induce practised and reliable *S*'s to judge "one" when two points were given and "two" when one point was given, and with considerable uniformity.[1]

(8) *Dependence on circulation of the blood.* The condition of the circulation in the region under test affects the limen. It appears from the studies of Brown-Sequard, Schmey, and others, that arterial hyperemia increases cutaneous sensitivity, whereas anemia, or venous hyperemia, or decided cold, reduces sensitivity. Excessive stretching of the skin decreases sensitivity, as does the use of narcotics.

(9) *Corrrelation with intelligence.* Schuyten seems to be the only investigator to discover a distinct positive correlation with mental ability. His very rapid tests (one per minute) on the cheeks of Antwerp pupils led him to conclude that the intelligent group had a decidedly acuter average threshold than the unintelligent group, and again he claims by this method to have been able to detect abnormal children readily and to classify a group of 60 children approximately as to mental power. Binet (2) found his intelligent superior to his unintelligent group at the first trial, but the difference soon lessened as the boys became adapted to the test. On the other hand, the rather crude test of Wissler (63) revealed no correlation with the class standing of university students, and the more careful work of Burt with Oxford school children gave the non-significant correlations of .13 and —.06, with large probable errors.

---

[1]These observations, which might be repeated in many other fields, show how essential it is to work methodically and under constant conditions. The extent to which *S*'s discrimination is affected by his attitude toward the experiment, and by his manner of judging in general, has led Binet to declare (5) that the "compass test" measures "tactual intelligence" rather than the fineness of touch itself.

Van Biervliet (55) used the compass test to secure a measure of intelligence, not by the limen itself, but by its mean variation, on the assumption that this latter measure is, in almost any test, the real index of intelligent work. His figures give the 10 most intelligent of 300 university students an index of 17.7, and the 10 least intelligent an index of 27.6. His index, however, as Binet (6a) has pointed out, is a rather dubious device.

(10) As might be expected, the use of the test for the examination of *abnormal children,* criminals, truants, etc., is beset with difficulty, because it exacts prolonged, sustained attention and interest (Kellor, 29; Kelly, 30). Simon (48) was unable to test the lowest grade of children in a school for the feeble-minded, but obtained results from the less defective types which indicated that their sensitivity was less than that of normal children. Whether the difference is reducible to differences in cutaneous sensitivity itself, or to differences in ability to control the attention and understand directions, is not clear. Burt found one congenital imbecile with a very low limen.

(11) *Dependence on cultural status.* Burt (11, 119-120) has assembled the results obtained by Rivers upon the Todas and by McDougall upon Papuans and Dayaks, and by joining them with other data by McDougall and himself upon various social classes in England, and making certain allowances for differences in method, has come to the conclusion that "the groups would fall in an order showing a complete inverse correspondence with that of cultural development," *i.e.,* "the least intellectual group tend in average tactile discrimination to be the more acute; while among individuals of the same cultural class, any apparent positive correlation between tactile discrimination and intelligence is probably illusory and due to the interpretative quickness of the more intelligent."

(12) *Other correlations.* Burt's tests showed a coefficient of reliability in the neighborhood of .74, but those of Krüger and Spearman of only .42, and it was partly for this reason, perhaps, that these latter investigators were unable to establish any correlation between the limen for twoness and capacity

in pitch discrimination, in addition, in committing to memory or in the Ebbinghaus completion method.

REFERENCES

(1) H. Adsersen, Eine aesthesiometrische Untersuchung. *ZScGd*, 27: 1904, 540-3.

(2) A. Binet, Attention et adaptation. *AnPs*, 6: 1899 (1900), 248-404.

(3) A. Binet, (*a*) Un nouvel esthésiomètre. *AnPs*, 7: 1900 (1901), 231-9; (*b*) Technique de l'esthésiomètre, 240-8.

(4) A. Binet, La mesure de la sensibilité. *AnPs*, 9: 1902 (1903), 79-128. Also five other articles on esthesiometry, pp. 129-252.

(5) A. Binet, De la sensation à l'intelligence. *RPhF*, 56: 1903, 450-467, and 592-618.

(6) A. Binet, (*a*) A propos de la mesure de l'intelligence. *AnPs*, 11: 1905, 69-82. (*b*) Recherches sur la fatigue intellectuelle scolaire et la mesure qui peut en être faite au moyen de l'esthésiomètre. *Ibid.*, 1-37. (Also reported as Expériences sur la mesure de la fatigue intellectuelle scolaire au moyen du sens du toucher. *BuSocEtPsEnf*, 5: 1905. 628-632, 644-652.)

(7) B. Blazek, Ermüdungsmessen mit dem Federaesthesiometer an Schülern des Franz-Joseph-Gymnasiums in Lemberg. *ZPdPs*, 1: 1899, 311-325.

(8) T. Bolton, The reliability of certain methods for measuring the degree of fatigue in school children. *PsR*, 7: 1900, 136-7.

(9) T. Bolton, Ueber die Beziehungen zwischen Ermüdung, Raumsinn der Haut und Muskelleistung. *PsArb*, 4: 1902, 175-234.

(10) N. Bonoff, Étude médico-pédagogique sur l'esthésiométrie et la simulation a l'école. *InMagScHyg*, 4: 1907-8, 384-394.

(11) C. Burt, Experimental tests of general intelligence. *BrJPs*, 3: 1909, 94-177.

(12) C. Burt, Experimental tests of the higher mental processes and their relation to general intelligence. *JEPd*, 1: 1911, 93-112.

(13) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.

(14) Buzenet, Mesure de la sensibilité tactile. *BuSocEtPsEnf*, 5: 1905, 633-4.

(15) J. N. Czermak, (*a*) Beiträge zur Physiologie des Tastsinnes. *Sitzungsber. d. k. Akad. d. Wiss.* (Wien), math.-nat. Kl., Abt. 1, 15: 1855, 482-521. (*b*) Weitere Beiträge, etc. *Ibid.*, 17: 1855, 577-600.

(16) F. B. Dresslar, Studies in the psychology of touch. *AmJPs*, 6: 1894, 313-368.

(17) M. Foucault, L'illusion paradoxale et la seuil de Weber. Montpellier, 1910. Pp. 211. (Critical review by A. Binet. *AnPs*, 17: 1911. 420-7.)

(18) G. Germann, On the invalidity of the esthesiometric method as a measure of mental fatigue. *PsR*, 6: 1899, 599-605.

(19) D. Gineff, Prüfung der Methoden zur Messung geistiger Ermüdung. Zürich, 1899. Pp. 68.

(20) H. Griesbach, Energetik u. Hygiene des Nerven-Systems in der Schule. München u. Leipzig, 1895. Pp. 97. Also published as Ueber Beziehungen zwischen geistiger Ermüdung u. Empfindungsvermögen der Haut. *Arch. f. Hygiene*, 24: 124.

(21) H. Griesbach, Weitere Untersuchungen über Beziehungen zwischen geistiger Ermüdung u. Hautsensibilität. *InMagScHyg*, 1: 1905, 317-417.

(22) H. Griesbach, Hirnlokalisation und Ermüdung. (Schlussfolgerungen des Autors aus *ArGsPhg*, 131: 1910, in *InMagScHyg*, 6: 1910, 174.)

(23) G. S. Hall, Laura Bridgman. *Mind*, 4: 1879, 149-172, especially 160 ff.

(24) V. Henri, (*a*) Revue générale sur le sens du lieu de la peau. *AnPs*, 2: 1895 (1896), 295-362 (with bibliography). (*b*) Ueber die Raumwahrnehmungen des Tastsinnes. Berlin, 1898. Pp. 228. (With extended bibliography.)

(25) V. Henri and G. Tawney, Ueber die Trugwahrnehmung zweier Punkte bei der Berührung eines Punktes der Haut. *PhSd*, 11: 1895, 394-405.

(26) D. S. Hill, Tests with a modified Binet-Buzenet esthesiometer. *PsBu*, 7: 1910, 66-7.

(27) J. Jastrow, Psychological notes on Helen Keller. *PsR*, 1: 1894, 356-362.

(28) C. Judd, Ueber Raumwahrnehmungen im Gebiete des Tastsinnes. *PhSd*, 12: 1896, 409-464. (Historical survey, 451-8.)

(29) Frances Kellor, Experimental sociology. N. Y., 1901. Pp. 316.

(30) R. L. Kelly, Psychophysical tests of normal and abnormal children; a comparative study. *PsR*, 10: 1903, 345-372.

(31) E. Kraepelin, Ueber Ermüdungsmessungen. *ArGsPs*, 1: 1903, 9-30.

(32) W. Krohn, Sensation-areas and movement. *PsR*, 1: 1894, 280-1.

(33) F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. *ZPs*, 44: 1906, 50-114.

(34) J. Leuba, On the validity of the Griesbach method of determining fatigue. *PsR*, 6: 1899, 573-598.

(35) Gladys W. Martin, The evidence of mental fatigue during school hours. *JEPd*, 1: 1911, 39-45, 137-147.

(36) Gladys W. Martin, A study of mental fatigue. *BrJPs*, 5: 1913, 427-446.

(37) E. Meumann, Vorlesungen zur Einführung in die exp. Pädagogik. 1st ed., Leipzig, 1907.

(38) P. Nolkow, Aesthesiometrische Ermüdungsmessungen. *InMagSc Hyg*, 4: 1907-8, 437-481.

(39) M. Offner, Mental fatigue. Eng. trans. Baltimore, 1911. Pp. 133.

(40) C. Ritter, Ermüdungsmessungen. *ZPs*, 24: 1900, 401-444.

(41) W. H. R. Rivers, Observations on the senses of the Todas. *BrJPs*, 1: 1905, 321-396, especially 363-371.

(42) Y. Sakaki, Ermüdungsmessungen in vier japanischen Schulen. *InMagScHyg*, 1: 1905, 53-100.

(43) Schmey, Ueber Modificationen der Tastempfindung. *DuBois-Reymond's Archiv*. (Physiol.), 1884, 309-312.

(44) R. Schulze, Aus der Werkstatt d. exp. Psych. u. Pädagogik. Leipzig, 1909. Pp. 292.

(45) M.-C. Schuyten, Sur les méthodes de mensuration de la fatigue des écoliers. *ArPs(f)*, 2: 1903, 321-6.

(46) M.-C. Schuyten, Comment doit-on mesurer la fatigue des écoliers. *ArPs(f)*, 4: 1904, 113-128.

(47) M.-C. Schuyten, Researches in the mental classification of children. (Summary from *PdWJb*, 7: 1909, 73-118, in *InMagScHyg*, 6: 1910, 179-180.) (See also Revue de psychiatrie, 1908, p. 135.)

(48) T. Simon, L'interpretation des sensations tactiles chez les enfants arriéres. *AnPs*, 7: 1900 (1901), 536-558.

(49) L. Solomons, Discrimination in cutaneous sensations. *PsR*, 4: 1897, 246-250.

(50) R. Sommer, Ausstellung v. exp. psych. App. u. Methoden. 1904.
(51) G. Tawney, The perception of two points not the space-threshold. *PsR*, 2: 1895, 585-593.
(52) G. Tawney, Ueber die Wahrnehmung zweier Punkte mittelst des Tastsinnes, mit Rücksicht auf die Frage der Uebung. *PhSd*, 13: 1897, 163-222.
(53) Helen B. Thompson, The mental traits of sex. Chicago, 1903. Pp. 188.
(54) J. van Biervliet, L'asymétrie sensorielle. *BuAcRoySci*, 34: Série 3, 1897, 326-366.
(55) J. van Biervliet, La mesure de l'intelligence. *JPsPa*, 1: 1904, 225-235.
(56) T. Vannod, (*a*) La fatigue intellectuelle et son influence sur la sensibilité cutanée. *RMdSuisse*, 17: 1897, 21. (*b*) La méthode esthésiométrique pour la mensuration de la fatigue intellectuelle. *Rep. 1st Intern. Cong. on Sch. Hygiene*, Nürnberg, 1904, vol. ii.
(57) K. Vierordt, Ueber die Ursache verschiedenen Entwickelung des Ortssinnes der Haut. *ArGsPhg*, 2: 1869, 298-306. Also Die Abhängigkeit der Ausbildung des Raumsinnes der Haut v. d. Beweglichkeit der Körperteile. *ZBi*, 6: 1870.
(58) A. Volkmann, Ueber den Einfluss der Uebung auf das Erkennen räumlicher Distanzen. *Ber. d. Sächs.-Ges. d. Wiss.*, 10: 1858, *math. phys. Abth.*, 38-69.
(59) L. Wagner, Unterricht u. Ermüdungsmessungen an Schülern des neuen Gymnasiums in Darmstadt. *SmAbPdPs*, 1: 1898, Hft. 4, Pp. 134.
(60) Margaret Washburn, Some apparatus for cutaneous stimulation (with cut). *AmJPs*, 6: 1894, 422-3.
(61) Margaret Washburn, Ueber den Einfluss der Gesichtsassociationen auf die Raumwahrnehmungen der Haut. *PhSd*, 11: 1895, 190-225.
(62) E. H. Weber, De pulsu, resorptione, auditu et tactu. Leipzig, 1834.
(63) C. Wissler, The correlation of mental and physical tests. *PsMon*, 3: No. 6, 1901. Pp. 62.

CHAPTER VII

Tests of Attention and Perception

The tests included in this chapter are those commonly assumed to measure such capacities as "power of observation," "quickness of perception," "range of attention," "mental grasp," etc. They are practically confined to the sphere of visual perception, and imply that this perception takes place under active attention: thus, in general, they seek to determine the subject's capacity to perceive visual objects or symbols when the conditions of perception are limited by short temporal persistence of the stimulus, or by other difficulties or complications that are intentionally introduced.

It has been said that experimental psychology discovered attention. Whether this be strictly true or not, every psychological experiment of necessity takes account of attention. And so, in every mental test that presupposes effort or concentration, we measure the capacity under investigation, always as conditioned by the particular degree of attention manifested at the time. It follows that a fundamental presupposition for the comparison of the results of such tests is that they shall all be secured under the same condition of attention. In practise, we find that the best way to secure this constant degree of attention is always to exact the maximal degree. Yet, in so far as the capacity to attend does differ in different individuals and in the same individuals at different times, just so far our tests of various other capacities, such as discrimination, retention, and the like, are often felt to be measures of attention, quite as much as measures of these other capacities.

Despite this fact, or perhaps on account of it, a direct measure or test of degree of attention is difficult to secure. In theory, since attention is a condition of consciousness in which certain constituent processes are clear and prominent, attention is directly measurable in terms of clearness. In practise, we must, in all probability, content ourselves with an attempt to measure attention indirectly, not by any single test, but by

a series of tests, all of which exact maximal effort. Even then, it must be recognized that we measure, not the process or condition of attention itself, but a product or concomitant of that condition.

A fundamental source of difficulty in these tests of attention lies in the fact that, with repetition, or even with the progress of the first trial, the task assigned becomes progressively easier; a tendency toward automatism appears, and the tax on attention diminishes proportionately.

TEST 24

Range of visual attention.—In a single 'pulse' of attention only a small number of impressions can stand out clearly: the area of span of consciousness is definitely limited. In the sphere of vision, we find that if we give but a single glance at any heterogeneous collection of objects, such as the goods displayed in a store-window, or the jumble of odds and ends in an old tool-chest, we are able to grasp and enumerate only a very few, perhaps four or five, of these objects.

For the scientific study of the area or range of visual attention, psychologists employ some form of short exposure apparatus, or **tachistoscope.**[1] (Greek, *tachistos,* very rapid, and *skopein,* to view). The essential idea of a tachistoscope is to furnish a field upon which *S* may fixate his glance and attention, and to supplant this field for a brief instant by another which contains the test-material. There is, then, a pre-exposure field (which contains a fixation-mark), an exposure-field, and a post-exposure field. The contents of the exposure-field depend, of course, upon the object of the experiment. In the main, the tachistoscope has been most used for the experimental investigation of the process of reading, and, accordingly, with an exposure field containing printed texts, isolated words, nonsense syllables, single letters, etc., but it has also been used for determining the range of attention for the visual apprehension of groups of dots, lines, geometrical drawings, objects, colors, etc.

[1] The name *Tachistoskop* was first employed by Volkmann (27).

As Dodge has remarked, "no psychological instrument is subject to greater modification in response to special experimental conditions than exposure apparatus," and it may be added that in no other experiment are the results more evidently conditioned by the form of apparatus and type of procedure employed.[1]

Wundt has formulated the essentials of a good tachistoscope as follows:

(1) The exposure must be short enough to preclude eye-movements.

(2) The arrangement of the fixation-mark and of the stimulus must be such that all the constituents of the exposed object can be seen with at least approximately equal distinctness, *i. e.*, the exposure-field must coincide with the ocular field of direct vision.

(3) The exposure of all parts of the field should be simultaneous, or so nearly so that there shall be no noticeable time-differences in the illumination of the various regions.

(4) Retinal adaptation must be favorable, and sudden transitions from dark to light must be avoided.

(5) Persistent after-images must be avoided.

(6) The duration of the retinal excitation must be limited enough to preclude roving of attention over the exposure-field.

(7) A ready-signal must be given at an appropriate time before the exposure.

Further requirements given by Dodge are as follows:

(8) The relative illumination of the pre-exposure, exposure, and post-exposure fields should be capable of experimental modification.

(9) The exposure should be noiseless and free from distraction.

(10) It should be possible to arrange for monocular or binocular observation.

Apparatus.—Disc tachistoscope (Fig. 57).[2] Frosted tubular lamp, 16 C. P. Two 4-inch clamps. Blanks of cardboard 9 cm. square. Two complete sets of Willson's gummed black letters

---

[1]For the history of the development of the tachistoscope and for the various controversies that have been waged concerning the necessity for absolutely simultaneity of exposure of the entire field, the extent to which convergence and accommodation are controlled by different forms of fixation-point, the actual and the optimal duration of the retinal excitation set up by the exposure, and the optimal conditions of general and local adaptation, the reader may consult the various references at the end of the test, especially Binet, Cattell, Dodge, Erdmann and Dodge, Goldscheider and Müller, Becher, Wundt, Huey and Zeitler. A short summary of these issues was published by the author in the first edition of this work. It may be pointed out merely that the black pre- and post-exposure fields of the disc tachistoscope here prescribed afford the most brilliant possible exposures, with longest retinal effects, so that an exposure of 50 to 75 σ will give a well "cleared-up image" (to employ Dodge's term) and at the same time eliminate eye-movement.

[2]This instrument, devised by the author and employed by him in studying the effect of practise on the range of attention does not fulfill all of the requirements of the ideal tachistoscope, but it has the merit of being relatively inexpensive, simple in operation and construction, and of answering satisfactorily for comparative tests. It has also been designed with a view for use both for these short and for much longer exposures (Test 25).

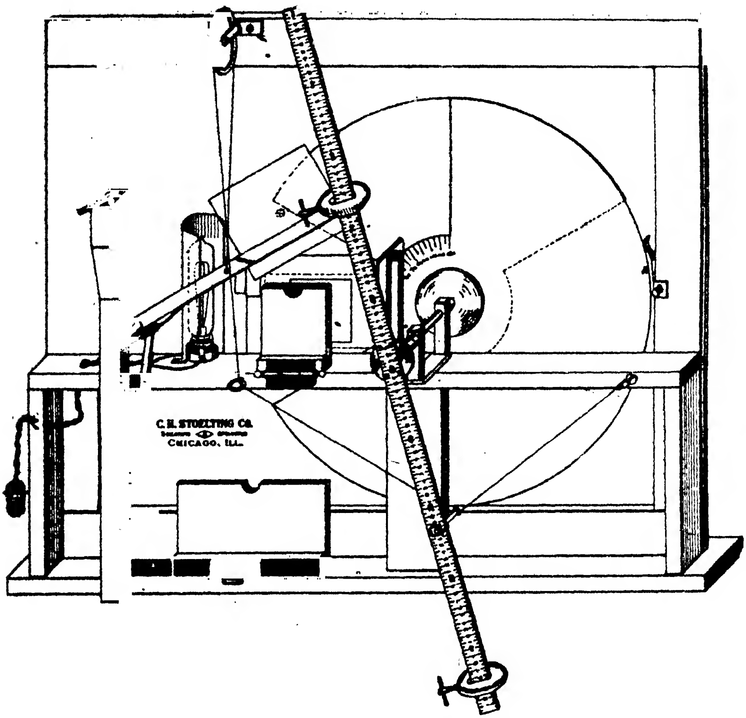


FIG. 57. DISC TACHISTOSCOPE.

and figures, Size 3. Drawing ink and ruling pen. Head-rest,[1] with suitable supports and clamps. [Fifty-vibrations dry-contact fork (Fig. 58.) Dry battery. Connecting wire.]

PRELIMINARIES.—Prepare a series of exposure-cards by use of the gummed letters and figures. Paste the letters smoothly and evenly, and center the series on each card. The test-objects should include isolated letters, groups of letters, or letters mixed with digits in nonsense arrangement, in numbers from two to eight or ten per card. Prepare other cards with short words, short sentences, or columns of digits. Prepare still others with the aid of drawing ink, so as to present regular or irregular series of spots, lines, geometrical figures, surfaces,

[1]Suggestions for the construction of head-rests are given by Judd.

etc., in varied fashion.[1] It is well to confine the objects, in so far as possible, to an area 50 mm. wide and 35 mm. high. Exposure cards can also be formed, after the method used by Huey, by pasting bits of printed texts of various dimensions on the blanks, or colored surfaces may be introduced *ad libitum*.

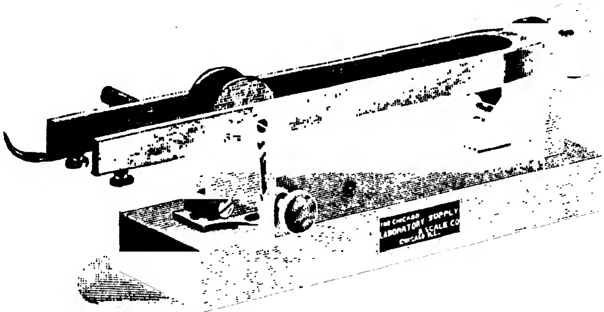


FIG. 58. FIFTY-VIBRATIONS DRY-CONTACT FORK.

Set up the tachistoscope, and clamp or screw it firmly to the table. Connect the electric light to a circuit of proper voltage. By means of the supports and clamps, adjust the head-rest so that *S*'s eyes shall be about 40 cm. in front of, and slightly higher than, the exposure card.

Adjust the tachistoscope for a point-exposure of 60σ.[2]

The pendulum-arm is attached with its mm. scale in view at the back of the instrument, and with the zero end of its scale lying upon the release-lever when the instrument is ready for exposure. To secure an exposure of 60σ, *E* should set the first weight at 5 cm. (measured at the edge of the weight nearer the lever), should set the second weight at 60 cm. and should open the sector 25 deg. He may fasten the overlapping edges of sector and disc by pushing a small paper-clip over them at the periphery. The entire disc is to be tightened up upon its axis in such a position as to bring the square opening in the disc in alignment with the

[1]For general suggestions, see Huey, 18, pp. 75 ff, and Binet, 3, p. 349. For special suggestions on the use of spots to test ability to apprehend number, see Freeman. A number of prepared cards accompany the set of material supplied by Rossolimo for his 'profile' method.

[2]The time of total exposure of several rows of objects is not the same as that of the exposure of any single point. If t is the time required for the notched portion of the disc to pass a given point on the exposure field, and L is the time of exposure of the group of objects to be determined, and if H and h are the heights of the total exposure field and of the portion of the field occupied by the objects, respectively, then $L = t\,(1 + \frac{h}{H})$.

square opening in the screen.[1] The axle must be properly oiled and its bearings in good alignment.

See that the device that lifts the fixation-card works smoothly and quietly. The movement of this card should be invisible to *S*, and the stimulus-object should be entirely unobstructed by it at the moment of exposure.

METHOD.—Seat *S* so that his head is supported in the head-rest without undue strain (Fig. 59). Set the tachistoscope for an exposure, and place a very simple exposure-card in the holder. In this position of the disc, the exposure-card is hidden by the fixation-card, which is visible through an opening in the disc. Instruct *S* to fixate the cross of the fixation-card as accurately and intently as possible, and, when he feels that his attention is thoroughly prepared, to signal for the release of the disc.[2] Let him then report, orally or preferably by drawing, what he has seen. Unless attention is manifestly poor, do not repeat the exposure.[3]

---

[1]If he wishes to measure the time of exposure accurately, *E* may attach a piece of smoked paper temporarily to the back face of the disc, connect the 50-vibs. fork with the battery, and adjust the fork so that its recording point (or the recorder of a sensitive signal-magnet, *e. g.*, a Deprez signal with which the fork is electrically connected) leaves its curve traced upon the paper when the disc is released. He may then determine the time of exposure by counting the number of 'waves' recorded while the notched portion of the disc is passing the center of the exposure field.

Or he may apply the fork directly to a smoked kymograph drum (see Test 10) and record the duration of the exposure with a signal-magnet on a parallel tracing. For this purpose, the two clips fitted with light connecting wire, with which the instrument is provided, are placed on the periphery of the disc in such a manner that, at the moment of exposure and at the moment of occlusion of the center of the field, they make electrical contact with the copper brush which is fastened to the frame of the instrument.

[2]The rather unusual procedure of allowing *S* to control initiation of the experiment is justified here by the fact that the signal for the release may be very simple, and that we are more likely to secure maximal attention from *S*. In fact, a simple mechanical release may be arranged for the disc tachistoscope by running a cord from the release-lever through two screw-eyes to a small lever convenient to *S*, who can then set off the exposure himself when he is quite ready. If, after several practise trials, it is evident that *S* is distracted by this procedure, *E* should revert to the usual method of giving the ready signal himself. The practise series, in any event, need be no longer than is required to accustom *S* to the general setting of the test.

[3]There is, however, good precedent for using repeated exposures, if desired: Cattell used a series of 5 exposures; Huey occasionally, and Binet regularly, made as many as 20 exposures; Titchener prescribes an indefinite number of exposures. This procedure is based on the assumption that no more is actually *seen* in 20 exposures than in 1, but that the series of exposures determines the limits of assimilative capacity. This method seems unnecessary, especially for comparative purposes.

Introduce more and more complex cards of the type under investigation until a limit is reached beyond which *S* cannot carry his observation. For the experiment proper, use in the main simple series of consonants, and drawings, but unless there is some special reason for it, do not change without notice from the one type of exposure-card to the other.

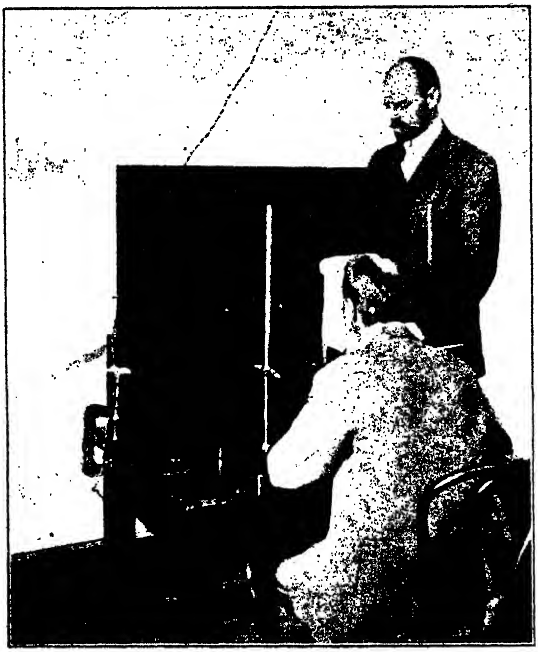


FIG. 59. TESTING THE RANGE OF ATTENTION.

VARIATIONS OF METHOD.—Test the range of attention with other forms of material, *e.g.*, digits, nonsense syllables, words, and especially with short sentences and drawings. Test the assimilative completion of word skeletons (groups of characteristic letters, mutilated or misspelled words). Vary the time of exposure, especially toward the shorter times. Try changes in the backgrounds and fields by covering the screen and disc with white, gray and black papers in different combinations, and using gray or black blanks for exposure-cards.

Try the cumulative method of exposure. Try reading distances greater than 40 cm. Try various sizes, forms, or colors of type.

TREATMENT OF RESULTS.—*S*'s rank is measured in general by the number of concrete objects correctly reproduced. This rank should be computed separately for different forms of exposure-card (if other than isolated letter tests are used) and for letter-series of different lengths, *e.g.*, for 6-place, 7-place, 8-place series, etc. In the case of letters or words, the simplest method of ranking is to assign one unit for each letter or word correctly reproduced, but to deduct 0.5 for errors of insertion or transposition. Thus, if the stimulus-card was FRMUTH, then FRMTUTH or FMRUTH would be ranked as 5.5, whereas FRMUT or FRMUH would be ranked 5. If desired, double errors, *e.g.*, transposition and insertion, as FMRTUTH, may be doubly discounted, but it is simpler to count this rendering also as 5.5.

In the case of more complex exposure-cards, such as drawings, it is sometimes possible to rank *S* on the basis of the number of lines reproduced, discounting 0.5 for lines incorrectly placed with respect to the total figure; in other cases, it is more satisfactory to assign a subjective estimate of the general fidelity of the reproduction on some arbitrary scale, *e.g.*, 10 for a perfect reproduction, 0 for absolutely nothing, and intermediate ranks in proportion.

If the method of serial exposures is followed, *S*'s rank can often be indicated by the number of exposures necessary to obtain accurate reproduction.

RESULTS AND CONCLUSIONS.—(1) When a series of *unrelated objects* is exposed, the average number of impressions that can be grasped in a single exposure lies between four and five. While, occasionally, *S*'s may grasp as many as seven impressions, this is usually due to a more or less well-recognized tendency to group or unify the objects in some manner. Table 32 summarizes the author's experiments (30) upon four college students with *letter-series,* 100σ exposure, and with the use of apparatus like that prescribed, save that the pre- and post-exposure fields were white. It will be seen that the average

TABLE 32.

Average Number of Letters Read Correctly in one Exposure (Whipple).

OBSERVERS	FIVE-PLACE SERIES	SIX-PLACE SERIES	SEVEN-PLACE SERIES	ALL
Mr. B	4.85	5.09	5.25	5.06
Mr. E	4.84	4.49	4.48	4.53
Mr. N	4.74	4.92	5.38	4.97
Mr. T	4.51	4.86	4.40	4.71
Average	4.74	4.84	4.88	4.82

performance for different *S*'s under different conditions is quite similar, and that it lies between four and five impressions. The 6- and 7-place series are somewhat more favorable because the *S*'s can occasionally reproduce 6 or 7 impressions.

These results are in close accord with those of other experimenters.[1] Cattell's tests (4) place the average limit for digits at 5, for letters at 3-4, and more often 3, though one *S* could grasp 6 letters. Erdmann and Dodge found that 6-7 letters could be read at times. Zeitler (30) points out that, while a series of consonants has an assimilative limit of 4-7, one may grasp 5-8 impressions, if vowels are interspersed. In the author's experiments, it was found, similarly, that the 7-place series that were read successfully were almost always those containing vowels which permitted the formation of nonsense syllables, *e.g.*, WAEGZME, KMDEMBH. The statement that the first and last letters are those usually seen clearly is not so easily confirmed and does not apply to the results of all *S*'s.

(2) *S*'s not infrequently report that more is seen than can be remembered a moment later when the report is given. In such cases, it is still often possible to state whether or not a given character was present.

(3) Despite the meaningless character of isolated letters, a series once exposed may be sufficiently well remembered to be

[1]The horizontal seriatim exposure used by Hylan is so little comparable with the whole-field exposure of others that his results may not be expected to conform with theirs. Hylan found that, with an exposure of 6 letters at the rate of 3.6 σ per letter, an average of 1.9 letters was read, while a longer exposure, 42 σ for the entire card, permitted 2.6 letters to be read.

recognized if used in the experiment again, even after a lapse of several days.

(4) When familiar syllables are combined to form *nonsense words, e.g., lencurbilber,* 6-10 letters can be grasped in one exposure (Zeitler).

(5 In the reading of *isolated words, i.e.,* collocations of words that do not make sense, the results of different investigators show a lack of accordance. As in the case of letters, Cattell's results indicate a lower range than those of other investigators. Cattell placed the limit of grasp at 2-3 short one-syllabled words. Erdmann and Dodge (8) found that, in a single exposure, 4 isolated words can almost always be read, and very often 5. Again, single words of 19-22 letters can be read in one exposure without roving of attention. Some of Zeitler's *S*'s could even read a 25-letter word, such as *Aufmerksamkeitsschwankung.* Becher read 26-letter words with a single, electric spark exposure. In general, the difficulty of grasping words does not increase in proportion to their length, while Cattell's reaction experiments showed that short words can be named more quickly than letters.

(6) When short words are combined into *simple sentences,* it is found that the average reading capacity is 4-6 words. Zeitler's *S*'s read such sentences when the total number of letters was 20-30; Erdmann and Dodge report 4-6 words of 2-10 letters each; Cattell found the average amount 4 words, occasionally 6, though *S* could read at times a 7-word sentence. One of the longest correctly read sentences contained 34 letters: *"Eine Tochter muss ihrem Vater gehorchen."*

Huey (17, 18) exposed for 15 σ lines cut from magazines. His results show clearly the striking individual differences in the reading range, which practically every investigator has noted;[1] thus, one *S* read on the average continuously and correctly 10.25 mm. of the line, another 21.33 mm., a third 23.80 mm. and a fourth 32.40 mm.[2] Under very favorable conditions (atten-

---

[1]Cattell concludes that there is a decided difference in the sensitiveness of the retinas in different *S*'s, but it is quite as likely that these individual differences are as much central as peripheral.

[2]In the type used, 40 mm. was equal approximately to 28 letter-spaces.

tion, subject-matter, etc.), one of Huey's *S*'s could read a stretch of line 50 mm. long (about half a line), *e.g.*, such phrases as "condition of consciousness," or "the whole body converges," but these are not ordinary performances.

In exposures of this sort the amount read to the left and to the right of the fixation-point is not at all equal, but varies in either direction according to the subject-matter, *e.g.*, in the second phrase above, the fixation-point was at the *o* of *whole*. The tendency is, as might be anticipated, to read more to the right than to the left. "In almost every case in which a large amount is read, far more is read to the right of the fixation-point than to the left."

In these, as in other exposures, the extent of reading is curtailed in proportion as the word-groups resemble isolated words, as when divided by punctuation-marks.

There is evidence to show that the unit of reading is the word, inasmuch as *S*'s see, or at least report, words or phrases rather than letters, even at the ends of the sections read.

(7) Freeman (11) made comparative tests of the ability of adults and of children aged 6 to 14 years to apprehend the number of objects (*spots* of light for the most part) thrown on a screen by a tachistoscope. The actual scope of attention, as estimated with the aid of introspection, was found not to differ much with age. The range averaged between 4 and 5 impressions for adults, while for children of 8 to 10 years judgments of number averaged only 4.5 per cent. poorer than those of adults when the number of objects was four or less. But for five objects children's judgments are 22 per cent. poorer than the judgments of adults—a difference due primarily to the superior ability of adults to form groups. "Not only do adults apprehend objects which are objectively grouped better than children, but they also have an advantage in dealing with objectively ungrouped objects which are beyond the scope of attention by means of subjective grouping."

Virtually the same results have been found with exposure-cards composed of series of 4 to 15 ruled perpendicular *lines*, 2 mm. apart. Cattell found that *S*'s could give the correct number of lines exposed, up to 4-6 only. Similarly, Goldscheider and Müller (12) tried various combinations of straight and curved lines (10 σ exposure), with the result that, if the arrangement was quite irregular, only 4-5 constituents could be grasped; but in proportion as the arrangement became more symmetrical (thus facilitating grouping or unitizing), the

number of constituents that could be grasped was increased. A symmetrical arrangement of simple perpendicular strokes increased the observation-limit to 7, while a combination of straight lines into squares, symmetrically arranged, permitted the apprehension of 5 squares, and hence of 20 constituent lines. Similarly, experiments with semicircles, ellipses, etc., confirmed the general principle that the number of constituent elements grasped in a short exposure is a function of the degree of combination which these elements permit.[1]

(8) The apprehension of *simple geometrical forms, e.g.*, circles, diamonds, oblongs, etc., cut from black paper and pasted on a white background, is easier than that of letters. Hylan's results, compared with his results for letters above, may serve for illustration. Six *S*'s averaged 7 forms; his poorest *S* averaged 4.5, and his best *S*, 9.5 forms.[2] Quantz's test of requiring *S*'s to name aloud, in order and as rapidly as possible, a series of geometrical forms, colors, or words, during an exposure of 0.5 sec. or 1.0 sec. is not strictly comparable to the short exposure tests: he found that forms could be named less rapidly than colors or words, and that, so far as forms are concerned, as many can be named with 0.5 sec. as with 1.0 sec. exposure (2.75 and 2.8, respectively).[3]

If *complex drawings* which are not clearly related to well-known geometrical figures are used, the test becomes more difficult because the visual image cannot be identified or held by the assistance of verbal associates (Binet, 3).

(9) *Practise* has a curiously small effect upon the range of attention, when once the period of preliminary habituation to the arrangement of apparatus and method is passed (Cattell, 4; Hylan, 20, p. 396). The chief feature of whatever practise

[1]An obvious illustration of this principle is seen in the reading of letters and digits themselves. For an account of Goldscheider and Müller's tests of the constituents of digits, etc., consult Huey (19, pp. 78 f).

[2]Some idea of the qualitative factors which influence the perception of liminal visual forms may be obtained from the experiments of Miss Hempstead, though these were conducted with different apparatus and by a different method.

[3]Huey's statement (19, p. 54) that, according to Quantz, more could be read in a short than in a long exposure, is not substantiated by Quantz's tables, which merely show that *relatively* more can be read in a half-second than in a second exposure.

can be detected is an increase in ability to group isolated impressions into combinations. "Practise tends to unite into a closer perceptive unity impressions first combined with difficulty" (Hylan). The practise effect for isolated letter series for a period of seven to ten days as found in the author's tests (30) is indicated in Table 33.[1] It is evident that, if we discount the improvement due to adaptation, there is but a small enlargement of the range through practise.

TABLE 33

Effect of Practise upon the Perception of Letters (Whipple)

	FIVE-PLACE SERIES			SIX-PLACE SERIES				SEVEN-PLACE SERIES		
OBSERVER	B.	N.	T.	B.	E.	N.	T.	E.	N.	T.
First period	4.87	4.44	4.50	5.03	4.75	4.38	4.73	4.25	4.90	3.83
Later period	—	4.87	—	—	—	4.85	—	4.02	5.54	—
Last period	4.78	4.77	4.50	5.25	5.08	5.06	5.08	4.90	5.40	5.80

(10) Aside from the work of Freeman already cited, the relation of the visual range of attention to *age* has been studied carefully only by Griffing (13), but by a method so peculiar[2] as to make the applicability of his results to ordinary conditions rather dubious. Table 34 presents Griffing's results in terms of the total number of letters correctly read in a series of 10 exposures of 6 letters each. Griffing concludes from these data that the number of visual impressions perceived "is a function of individual growth, reaching its maximum only when the observer is fully developed."

[1]The four *S*'s did not take the same series, nor work for the same length of time, so that the data are too few to permit accurate averaging into three periods, save for *N*., and for *E*. in the 7-place series.

[2]Griffing's method was to expose with a fall-tachistoscope for 0.1 sec. or 1.0 sec., to a group of 10 to 20 *S*'s, six printed capital letters, 48 mm. high, but to vary the interval between the ready-signal and the exposure from 6 sec. to 1.5 min., or even 4 min., without the knowledge of the *S*'s, with the idea of testing "the observer's powers of prolonged attention" by keeping them waiting an indefinite time for the exposure. In view of the well-known irregularity and fluctuation of attention under such conditions, it seems clear that the degree of attention present at the moment of exposure, at least in a short series of tests, is almost a matter of chance.

TABLE 34

Relation of Visual Range of Attention to Age (Griffing)

NUMBER OF S S	AGE	LETTERS READ
39	7–9	4
77	10–12	13
73	13–15	18
132	16–18 +	27

(11) Griffing found no difference between the *sexes.*

(12) There is only questionable evidence of a relation between the *range of visual attention and mental ability.* Griffing divided his pupils into three groups on the basis of teachers' estimates, and found that his 'A' group had a somewhat higher average range of attention than the other groups, but that there were marked exceptions, so that "many pupils must have good powers of attention even when they show no evidence of them to their teachers." It is possible, however, that the outcome of Griffing's test is more dependent on the interest and good-will of the pupils than on their intelligence, for, as Griffing himself says: "Children of the most active minds would be most interested in novel experiences." Cattell states that in his tests "obtuse porters" required three times as long as educated persons to read a letter or word. Binet could not differentiate his bright from his dull children by the exposure of single words, but could differentiate them very clearly by the exposure of a drawing 20 times in succession.[1]

(13) The *qualitative analysis* of the perceptive processes concerned in reading during short exposures has developed differences of opinion with regard to the following points: (*a*) Is it possible for roving of attention to occur in exposures which are sufficiently short to eliminate eye-movement? (*b*) Do we apprehend words by wholes or by parts? (*c*) Are there

[1]Binet's *S*'s drew what they could of the drawing after each exposure of 70σ. An interval of 5-10 sec. was interposed between the exposure and the drawing. Only three of 11 *S*'s were able to give a correct copy within the limit set to the test; these three *S*'s were all of the group selected for superior intelligence. For reproductions of the actual drawings, consult Binet, 3, pp. 351-360.

certain letters or combinations of letters which give the cue for the perception of words, and if so, what are these letters? What share does the total length and general visual contour of a printed word have in its perception? (*d*) Do different readers adopt different methods of reading?

Without going into details[1] it may be said that the seeming lack of agreement between the results of different investigators is due in part to the divergence in experimental conditions, in part to the divergence in type of the *S*'s employed. We do not always read by wholes: neither do we always read by successive apprehension. The unit of attention, of visual apprehension, in reading is a variable quantity. In normal reading, assuming equally easy subject-matter, the manner of reading will depend upon the type of the reader. Aside from differences in speed and fluency of reading, we may probably distinguish two fundamental types of readers, the subjective and the objective. The latter exhibit the following characteristics: in tachistoscopic tests, their attention is directed to the optical fixation-mark; their range is small, *e.g.*, three isolated letters or one 12-letter word, but they read accurately, are quite certain of what they do see, and seldom guess; for them there exists a distinct time-interval between the visual perception and assimilative interpretation or rise of meaning, and they seldom confuse these two phases. The subjective readers differ from the objective in every one of these points: they can read words lying in indirect vision; they have a range of five isolated letters or one 27-letter word; their attention is placed mainly on the interpretative or assimilative phase, and their reading occurs mainly by large word-wholes, or even by phrases, on the basis of relatively meager visual cues. These subjective readers are not, however, necessarily the faster.[2]

If, now, the subject-matter is difficult, the tendency for the subjective type of reader is toward the more extensive use of visual symbols, and here, doubtless, as in the case of short

[1] Readers who are particularly interested in these special problems will find a discussion of them in the first edition of this work, pp. 237-240, which is here omitted for want of space and because the subject-matter belongs more to the psychology of reading than to the psychology of tests.

[2] For a more detailed discussion of these types, see Meumann, ii, 250 ff.

exposures, dominating letters or complexes become more important: it will depend upon the difficulty of the passage whether these letters play their chief rôle as prominent elements in the configuration of the word, or whether they are directly the object of attention.

If the subject-matter is very difficult, *e.g.*, the reading of beginners, or of an adult in a foreign language, especially if reading unfamiliar characters, such as Greek or Hebrew, we have an extreme case in which reading may, and often does, proceed letter by letter.

The reading of children, once the primary mechanical difficulties are mastered, is almost always of the subjective or interpretative type (Messmer), though for different reasons than in the case of the subjective reading of adults. That, however, there are many exceptions to this generalization is shown by the recent work of Pintner (25a).

REFERENCES

(1) N. Baxt, Ueber d. Zeit welche nöthig ist, damit ein Gesichtseindruck zum Bewusstsein kommt, etc. *ArGsPhg*, 4: 1871, 325-336.

(2) E. Becher, Experimentelle u. kritische Beiträge zur Psychologie des Lesens bei kurzen Expositionzeiten. *ZPs*, 36: 1904, 19-73.

(3) A. Binet, Attention et adaptation. *AnPs*, 6: 1899 (1900), 248-404.

(4) J. Cattell, The inertia of the eye and brain. *Brain*, 8: 1886, 295-312; also as Ueber die Trägheit der Netzhaut u. des Sehcentrums. *PhSd*, 3: 1886, 94-127.

(5) R. Dodge, An improved exposure apparatus. *PsBu*, 4: January, 1907, 10-13.

(6) R. Dodge, The psychology of reading. *PsR*, 8: 1901, 56-60.

(7) R. Dodge, An experimental study of visual fixation. *PsMon*, 8: No. 4, 1907, (No. 35). Pp. 95.

(8) B. Erdmann and R. Dodge, Psychologische Untersuchungen über das Lesen. Halle, 1898. Pp. 360.

(9) B. Erdmann and R. Dodge, Zur Erläuterung unserer tachistoskopischen Versuche. *ZPs*, 22: 1900, 241-267.

(10) S. Exner, Ueber die zu einer Gesichtswahrnehmung nöthige Zeit. *Sitz. Ber. d. Wiener Akad.*, 58: (2), 1868, 601-632.

(11) F. N. Freeman, Experiments on the perception of number in children and adults. *PsBu*, 8: 1911, 43-44.

(12) A. Goldscheider and R. Müller, Zur Phys. u. Path. des Lesens. *Zeits. f. klin. Med.*, 23: 1893, 131-167. For critical summary, see Wallaschek, *ZPs*, 7: 1894, 228-231.

(13) H. Griffing, On the development of visual perception and attention. *AmJPs*, 7: 1896, 227-236.

(14) F. Hamilton, The perceptual processes in reading. *ArPs(e)*, No. 9, December, 1907 (Columbia Univ. Contr., etc., 17: No. 1).

(15) H. Helmholtz, Handbuch d. physiol. Optik. Hamburg and Leipzig, 1896.

(16) Louise Hempstead, The perception of visual form. *AmJPs*, 12: 1901, 185-192.

(17) E. B. Huey, Preliminary experiments in the physiology and psychology of reading. *AmJPs*, 9: 1898, 575-586.

(18) E. B. Huey, The psychology and physiology of reading. *AmJPs*, 11: 1900, 283-302 and 12: 1901, 292-312.
(19) E. B. Huey, The psychology and pedagogy of reading. New York, 1908. Pp. 445.
(20) J. P. Hylan, The distribution of attention. *PsR*, 10: 1903, 373-403 and 498-533.
(21) C. Judd, (*a*) Laboratory manual of psychology. N. Y., 1907. Pp. 124. (*b*) Laboratory equipment for psychological experiments. N. Y., 1907. Pp. 257.
(22) O. Messmer, Zur Psychologie des Lesens bei Kindern u. Erwachsenen. *ArGsPs*, 2: 1903, 190-298.
(23) E. Meumann, Vorlesungen zur Einführung in die exp. Pädagogik u. ihre psychol. Grundlagen. 2 vols. Leipzig, 1907. Pp. 555 and 446.
(24) A. Michotte, Description et fonctionnement d'un nouveau tachistoscope de comparaison. *ArPs(f)*, 12: 1912, 1-13.
(25) W. B. Pillsbury, Attention. London and N. Y., 1908. Pp. 330.
(25a) R. Pintner, Untersuchungen über die Aufmerksamkeitsformen der Kinder beim Lesen und Reagieren. *PdPsArb*, 4: 1913, 1-54.
(26) J. Quantz, Problems in the psychology of reading. *PsMon*, 2: No. 1, 1897.
(27) E. C. Sanford, The relative legibility of the small letters. *AmJPs*, 1: 1888, 402-435.
(28) E. B. Titchener, Experimental psychology, vol. 1, Qualitative experiments. N. Y., 1901.
(29) A. Volkmann, Das Tachistoskop, etc. *Sitz. d. k. s. Ges. d. Wiss. Leipzig* (*math.-phys. Cl*), 1859, 90 ff.
(30) G. M. Whipple, The effect of practise upon the range of visual attention and of visual apprehension. *JEdPs*, 1: 1910, 249-262.
(31) F. Wirth, Das Spiegeltachistoskop. *PhSd*, 18: 1903, 687-700.
(32) W. Wundt, Zur Kritik tachistoskopischer Versuche. *PhSd*, 15: 1899, 287-317, and 16: 1900, 61-70. See also his *Völkerpsychologie*, vol. 1, Die Sprache, 1900, 530 ff, and his *Grundzüge d. physiol. Psychologie*, 5th ed., 1903, 3: 611 f.
(33) J. Zeitler, Tachistoskopische Versuche über das Lesen. *PhSd*, 16: 1900, 380-463.

TEST 25

Visual apprehension.—This test is, in many respects, similar to that of the range of visual attention (Test 24), but it differs in certain important particulars: the period of exposure is increased from a small fraction of a second to several seconds, and the test-object is correspondingly more complex. Both eye-movement and roving of attention take place, so that we cannot speak of the range of attention;[1] we are measuring, rather, the capacity to apprehend a number of disparate objects by visual examination during a short period.

Tests of this type have been variously designated as tests of "quick perception," of "observation," of "degree of attention,"

[1]If one accepts the view of Pillsbury and Hylan that attention in tachistoscopy is really successive, this test differs from the former only in degree: we now give opportunity for a greater number of successive acts of attention.

or even as "memory tests." The term "apprehension," however, seems suited to describe the psychological processes under examination, though it is impossible to draw any hard and fast lines between tests of apprehension, tests of memory and tests of fidelity of report.

This test has long been used as a source of amusement in competitive parlor games, and it has been urged by some writers, particularly by Miss Aiken (1), as a source of mental training[1] in the form of systematic schoolroom exercises. Miss Aiken's suggestions have been subjected to experimental test in the laboratory by the author and by Foster and in the schoolroom by Dallenbach. These investigations, aside from the work of Quantz (9) in his investigation of the psychology of reading, and of Jones (6) in a study of individual differences of children, seem to be the chief instances of the study of visual apprehension under experimental conditions.

Two forms of test are described: a special variation of one form is presented in Test 25A; other variations may easily be worked out to suit special conditions.

A. THREE SECONDS' EXPOSURE WITH THE TACHISTOSCOPE

APPARATUS.—Disc tachistoscope[2] and other material as in Test 24, save that the cardboard blanks are replaced by blanks of stiff paper 12.5 × 20 cm., and that the exposure-card holder

[1] The famous conjurer, Robert Houdin (4) used what he termed "perception by appreciation" as a basis for certain feats of "second sight." The capacity which he and his son Emile attained is so marvellous as to be worthy of record. For preliminary tests Houdin tried the estimation of the numbers of dots on dominoes till, he says, "we at length were enabled to give instantaneously the product [sum?] of a dozen dominoes. This result obtained, we applied ourselves to a far more difficult task, over which we spent a month. My son and I passed rapidly before a toy-shop, or any other displaying a variety of wares, and cast an attentive glance upon it. A few steps further on we drew paper and pencil from our pockets, and tried which could describe the greater number of objects seen in passing. I must own that my son reached a perfection far greater than mine, for he could often write down *forty* objects, while I could scarce reach thirty. Often feeling vexed at this defeat, I would return to the shop and verify his statement, but he rarely made a mistake."

This capacity may be compared with the results obtained by adults after practise for a month in the author's experiments: here the average number of objects named was but *six*, the maximum ten.

[2] A simpler, but less efficient, substitute may be contrived after the style of the 'krypteon' described by Sanford (p. 403).

and the opening in the screen of the instrument are adjusted to corresponding dimensions. *Collection of pictures of all kinds, cut from old magazines and trimmed to a size not over 12.5 × 20 cm. [Seconds' pendulum, Fig. 28, or metronome with electrical contacts (Fig. 21) and double time-marker, Fig. 27, in place of the 50-vd. fork.]

Preliminaries.—Set up the tachistoscope as described in Test 24. Remove the fixation-card, or turn it off to one side. Replace the small by the large card-holder, and arrange the screen so that the rectangular replaces the square opening. Adjust the disc so that the pre-exposure section just covers the exposure-card when the instrument is set.[1] Adjust the weights and the sector of the disc to give a point-exposure of 3 sec. This will be secured approximately by setting the first weight at 10 cm., the second at 85.2 cm., and opening the sector 115 deg.[2] For rough determination of the exposure, the stop-watch (Fig. 20) may be used. For accurate determination, use the seconds' pendulum or metronome and the signal-magnet (time-marker) mentioned in Test 10; make the determination as directed in Test 24.

The following *types of exposure cards* will be found desirable; 10-20 cards should be prepared of each type selected.[3] (*a*) Groups of irregularly arranged *dots* or small circles or crosses, in number from 6 to 20.[4] These should be grouped in a space not exceeding 35 mm. square in the center of the blank, and care must be taken to avoid too obvious hints of combinations in the groupings. These cards may readily be made by the use of the asterisk (*) sign of a typewriter. (*b*) Cards on which *pictures*—drawings, cuts, lithographs, etc., cut from magazines—have been pasted. (*c*) Cards containing a single line of 8-10 three-letter *nonsense syllables*. These may be prepared most easily by the typewriter. As it is important to

---

[1]Fixation is not so exact, of course, with this arrangement, but in view of the long exposures, the error may be neglected for comparative work.

[2]To get full 3 sec. exposure it is imperative that the bearings of the axle of the instrument be in perfect alignment and properly oiled.

[3]Consult Foster (pp. 13-14) and Dallenbach for suggestions on the preparation of various types of material.

[4]Cf. Aiken (1, p. 37), who terms this the test of "unconscious counting." Groups of about 14 dots prove satisfactory for testing adults.

avoid syllables that resemble words when pronounced, a selected list is given here. The last column may be used to illustrate the effect of using sense syllables. (*d*) Cards containing varied com-

tob	vap	urs	mib	kun	zib	gos	orl	ith	tas	arc
arg	ept	ibe	ong	pof	orm	eig	lin	enf	ech	ton
ime	ull	zin	acq	jek	ige	buh	spo	ipp	ume	add
arl	omb	irm	ruv	euf	gur	ite	pru	baw	ret	red
elt	ilt	smi	nen	gla	lud	rad	lel	cha	heb	eel
vid	zet	dak	ung	geg	tau	ahn	uff	teg	ruj	not
ool	euk	sef	ank	jur	rik	nuc	tef	lom	fid	rat
bli	bri	tud	ift	aum	yef	rin	orp	pud	gom	low
ild	rud	rab	urf	ked	geb	pum	gah	arb	zan	end
vel	zen	vem	eit	nis	zud	dro	wol	zig	luh	tan

binations of lines in meaningless *drawings*. Make these with pen and ink; keep the drawings within an area 75 mm. high and 130 mm. long; use combinations of arrows, circles, loops, and straight lines, with 10 to 15 of these elements in each drawing. One or two drawings with some hint of meaning, such as a conventionalized desk-telephone, a 'woodeny' disjointed horse, etc., may be introduced if *E* desires to secure light upon *S*'s use of verbal and other associations for holding the drawings in mind. (*e*) Cards containing typewritten four-line *stanzas* from some not too difficult nor too well-known poem. The object in using poetry is to secure a certain degree of equivalence in length, rhythm, style, and topic from one exposure-card to the next. (*f*) Cards containing columns of *digits*, as illustrated in Aiken, p.30. A sample column from her book is the following:

230
729
11
36
40000
16
40

METHOD.—Proceed in general as in Test 24 (save that there is no preliminary fixation-mark). Inform *S* what *type* of card is to be exposed, *i.e.*, a drawing, a stanza, etc.

S's reports may be as follows: for *dots*, a statement of their number, supplemented, if desired, by a rough pencil sketch of their position and an account of their grouping; for *nonsense syllables, poetry*, or *digits*, an oral or written report; for *drawings*, a pencil sketch; for *pictures*, a verbal description, supplemented, if *S* desires, by a pencil sketch. In the case of pictures, *E* may also quiz *S* with regard to the observation of various details which *S* has not reported, to see whether they have actually escaped his observation or have merely been neglected in his report, *e.g.*, by the use of such questions as: "Did you notice any details in the background?" "Is there any printing in the picture?" "What color was the girl's dress?" After *S*'s report is finished, *E* may often obtain further light upon *S*'s work by confronting him with the stimulus-card and asking him what features or details he had failed to note, what he had forgotten, or what he had misapprehended.

TREATMENT OF RESULTS.—Some of the material lends itself well, some but poorly, to quantitative treatment, but the latter is often most useful for qualitative analysis of the mental processes concerned. The following system, though obviously arbitrary, has been found serviceable. (*a*) *Counting dots:* assign credit for the reporting of the correct number only; consider any mistake, even of one number, a failure. (*b*) *Pictures:* estimate *S*'s grade upon a scale of 10, *i.e.*, assign the grade 10 to a report which seems to indicate complete recall of all the salient features of the picture; assign the grade 0 to complete failure, and score intermediate grades accordingly. This grading may be made fairly objective by counting up the number of features or 'points' which are adjudged essential to a satisfactory report, and comparing the number given by *S* with this standard number for the picture in question. (*c*) *Nonsense syllables:* assign one unit to each letter correctly reported (*e.g.*, 4 syllables = 12), but deduct 0.5 for each error of transposition or insertion, whether of letters within syllables, or the syllables themselves.[1] (*d*) *Drawings:* as in the case of pictures, rate the reproduction on a scale of 10, by reference to

[1]For a somewhat more elaborate, though still simple method of scoring nonsense syllables and words with reference to the correctness of their position as well as of their composition, see Lyon (7).

the number of lines or elements correctly reproduced, in comparison with a standard number for the drawing in question. (*e*) *Poetry:* assign one unit for each word correctly reproduced, but deduct 0.5 for each error of transposition or insertion. (*f*) *Digits:* as with nonsense syllables, assign one unit for each digit correctly reported, but deduct 0.5 for each error of transposition or insertion, whether of digits within numbers or of the numbers themselves.

B. SIX SECONDS' EXPOSURE WITHOUT THE TACHISTOSCOPE

MATERIAL.—Small table. Piece of cloth, preferably gray, large enough to cover the table-top. Seconds' pendulum (Fig. 28). A piece of cardboard about 30 × 45 cm., and a full-sized sheet (22 × 28 in.) of gray cardboard. Collection of miscellaneous small objects, familiar enough to be named by all the *S*'s, *e.g.*, pencil, rule, spoon, tin box, leaf, cup, bunch of keys, toy animals, salt-shaker, postcard, etc. Ten different objects will be needed for each exposure.

METHOD.—Place the large sheet of gray cardboard on the table. Arrange on this as a background a group of ten objects, but avoid combinations of obviously related objects, such as pen and inkstand. Make a rough sketch of the group, with a list of the objects, so that it can be restored when desired for later tests. Cover with the gray cloth (to conceal the objects while *S* is taking his position and receiving instructions). Let *S* now stand in front of, and close to, the table, but in a position that will not interfere with its full illumination. Inform him that he will be given 6 sec. to view a group of familiar objects, after which he will be asked to enumerate as many of these objects as possible and further to describe them briefly. Let him hold the smaller sheet of cardboard so as to cut off the view of the table-top. Start the seconds' pendulum, which must be placed somewhat to one side, where it can easily be seen by *E*, but will not distract *S*. Remove the cloth and take the cardboard which *S* has been holding. Give *S* a "ready" signal, and 2 sec. later, quickly remove the cardboard screen. At the expiration of 6 sec.[1] again cut off *S*'s view with the screen. Let

[1] The simplest method is to count the strokes mentally, "one, two, three," etc.: if the exposure is made at "one," the screen is to be restored, of course, at "seven."

him immediately turn his back to the table, and give a verbal description of the objects.

The chief stress should be placed on naming as many objects as possible: afterwards, *S* may be asked to describe the details of the objects or to make a rough sketch to indicate their relative positions. For qualitative purposes, *S* should be encouraged to give an account of the manner in which he observed the objects and the manner in which he has reproduced them. After *S* has enumerated as many objects as possible, exhibit the objects that were unnamed, either singly, or mingled with a number of objects not on the table, and ask *S* if he can identify any more of the objects exposed.

TREATMENT OF RESULTS.—Credit *S* with one unit for each object named.

RESULTS FOR BOTH METHODS.—(1) The author's experiments with adults brought out strikingly the very small increase in the *range of apprehension in comparison with the range of attention* (Test 24). Thus, in an exposure of 10-50σ, an average *S* can grasp 4 or 5 objects: here, with an exposure more than 100 times as long, the average *S* enumerates but 6 objects (with a mimimum of 3, and a maximum of 10). Similarly, 3 sec. exposure of nonsense syllables allows, on the average, 10.15 letters (Table 35), *i.e.,* between 3 and 4 syllables, to be read correctly, which is approximately the same as can be read with exposures of a small fraction of a second. On the other hand, the 3 sec. exposure of sense material gives an average range of nearly 12 words in contrast to the 4-6 word limit for ordinary tachistoscopy. This interesting advantage of sense material in the longer exposure is evidently due to the fact that such material can be grouped and recalled by larger and more meaningful units, whereas the heterogeneous combinations of nonsense syllables or disparate objects are more difficult to identify and recall. In the case of poetry, *S*'s feel that the limit of their performance is set simply by the amount that can be read during the exposure, whereas, even in the 6 sec. exposures, there is not time enough clearly to apprehend 10 disparate objects. The maximal reproduction of poetry with 3 sec. exposure was the first 19 words of the following:

"Were they unhappy then? It cannot be.
  Too many tears for lovers have been shed,
Too many sighs give we to them in fee,
  Too much of pity after they are dead."

(2) *Individual differences* in capacity for quick apprehension are clearly indicated by Table 35. Thus, *V* excels in the estimation of dots and in reading poetry, but is the poorest *S* in reading nonsense syllables, in reproducing drawings, or in describing pictures and objects. *G* excels in these performances, but is handicapped in reading poetry by his relative

TABLE 35

Individual Differences in Visual Apprehension (*Whipple*)

MATERIAL	DOTS*	PICTURES	NONSENSE	DRAWINGS	POETRY	OBJECTS
G	11/33	6.96	10.90	8.65	9.42	7.10
R	16/35	6.89	10.70	6.42	12.92	5.57
V	28/35	4.40	8.85	3.70	13.21	5.50
Average		6.09	10.15	6.26	11.83	6.03

*Efficiency is indicated here by the number of times the group of dots was correctly reported, *i. e.*, 11 out of 33 trials, etc. As *G.* missed two trials, the average cannot be figured exactly.

unfamiliarity with English poetry (he is of German descent): his poor capacity in estimating dots cannot be explained.

These results indicate that it is not possible to assert that an *S* has a given grade of general ability of apprehension, or even of visual apprehension: rather, we must state that he excels in the attentive observation of pictures, of drawings, of words, or of certain kinds of objects, etc. This confirms in an interesting way the general verdict of experimental work that mental ability is narrow and specific: here, for instance, we find that *V* is more than twice as efficient as *G* in the quick perception of groups of dots, while *G* is more than twice as efficient as *V* in the quick perception of irregular drawings.

These results obtained by the author are confirmed by subsequent investigators. Thus, Foster's three *S*'s, all graduate students in psychology, scored .97, .85 and .80 in the apprehension of nonsense syllables and .82, .65 and .74, respectively, in the apprehension of sense material. Dallenbach discovered decided differences in the capacities of individual school children in the same school grade, and Jones' tests with cards con-

taining 10 familiar objects gave an average reproduction of 4.98 objects,[1] but with extremes of 2.6 and 7.4 in average scores for trials with ten sets of objects exposed for 1.5 sec.

(3) *Dependence on sex.* Dallenbach's work with 2d grade pupils showed that boys surpassed girls in nearly all the tests. His supplementary trials with adults in university classes likewise indicated a superiority of men over women in visual apprehension.

(4) *Dependence on age.* Dallenbach found a direct, though small correlation with age in the work of the pupils in the 2d grade. University students without previous practise were found to surpass, on the average, the performance of the grade pupils who had had three months daily practise, though the two groups overlapped, so that the best group in the 2d grade surpassed the poorest group of university students under these conditions.

(5) *Dependence on practise.* The work of Foster and of the author was specially directed upon the possibility of improvement by practise in the case of adults, that of Dallenbach upon the possibility of improvement in the case of children. The results seem to indicate distinct differences in the practise possibilities of children and of adults. Thus, the results obtained by the author (Table 36), in which each 'period' represents the average of three exposures, usually one daily for three days, afford little warrant for the belief that systematic practise would enable an adult *S* markedly to improve his ability

TABLE 36

Effect of Practise upon Visual Apprehension. Average for Three Observers (Whipple)

PERIOD	1	2	3	4	5	6	7	8	9
Pictures	6.6	4.9	5.9	6.3	6.9	6.9	6.5	5.5	
Nonsense	9.3	10.6	8.4	10.8	11.7	10.6	10.4	8.9	9.2
Drawings	6.6	6.3	5.0	5.6	5.0	7.7	6.5	6.6	5.7
Poetry	10.7	11.5	11.3	10.8	13.0	12.5	13.0	11.7	
Objects	5.6	6.3	5.9	6.0	5.9	6.5	6.4		

[1]In the original article the average is erroneously given as 6.3.

for quick visual perception. The tests with dots do not lend themselves readily to quantitative treatment. The seeming improvement with drawings during the 6th period was due to the use of one very easy drawing in that group. There is some slight evidence of an improvement in reading poetry, which amounts roughly to the addition of one word, but this may be attributed to increased familiarity with the peculiar style of the poem in use. If any improvement can be inferred in the case of the objects-test, it must amount, on the average, to the addition of less than one object.

Foster sought to induce a direct improvement in the capacity for visualization itself. He found a net practise gain for three adults during 40 hours of practise extending over 10 weeks time of from 6 to 44 per cent., depending on the observer and on the material, but no part of this gain was ascribed to direct improvement of visualization.

The gain is, on the contrary, ascribed to these factors:
1. "Confidence and 'doing one's best' replaced discouragement and 'giving up.'
2. "Familiarity with material lessened the difficulty.
3. "The observers learned where and how to distribute attention effectively.
4. "More efficient methods of work were adopted. Tricks of counting, naming, grouping, etc., were discovered and used.
5. "Regular and definite procedure replaced hap-hazard, unorganized procedure."

Foster further found that practise-gains were large in the earlier and small in the later stages of the experiment and that they were much greater with nonsense than with sense material (28 vs. 10 per cent.).

Dallenbach's work with 5-sec. exposure of visual material before a class of children aged 7 to 10 years extended in daily exercises over a period of 17 weeks. Like Foster he found the improvement in visual apprehension rapid at first, then slow. When these pupils were divided into three groups on the basis of their ability first shown, the curve of improvement of the poorest group shows interesting divergences: these pupils had a slower initial rise in capacity, but eventually surpassed the group of medium ability (Fig. 60). When retested after 41 weeks of no drill, the children in the experiment showed them-

selves distinctly superior to classmates of their age who had not been through the drill, so that we here have apparent evidence of a permanent improvement in capacity effected by drill in visual apprehension.[1]

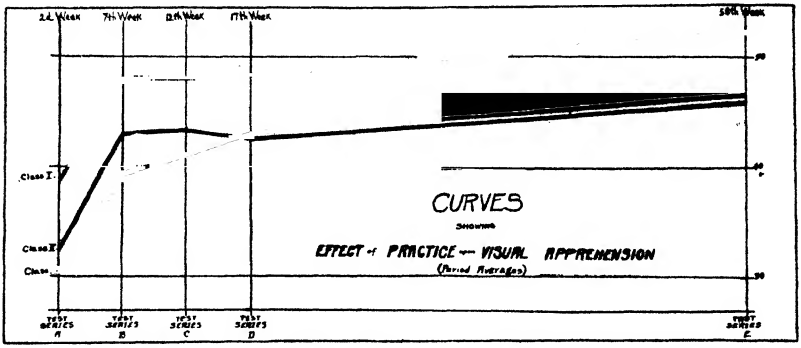


FIG. 60. EFFECT OF PRACTISE UPON VISUAL APPREHENSION (DALLENBACH).

(6) *Correlation with other abilities.* Dallenbach found a positive correlation between capacity for visual apprehension and school standing, amounting in one group to .85, in another to .20. Jones' tests gave a negative correlation, — 0.15, between visual apprehension of a group of 10 objects and auditory memory for words naming common objects.

(7) A *qualitative analysis* of the data secured in this experiment shows that the efficiency in visual apprehension exhibited by any *S* will depend on the following:

(*a*) *Native capacity* for concentrating attention in general. This is the factor which it is desired primarily to isolate and measure, but it is impossible to secure such a measurement until the other factors are measured or eliminated.

(*b*) The *degree of attention* given at the exposure in question. In theory, each exposure is accompanied by *S*'s maximal attention: in practise, this is not always secured. Tests in which *S* reports distraction must be thrown out. The effect of good preparedness on *S*'s part may be illustrated readily by exposing test-cards without warning of their type; the consequent elimination of 'expectant attention' will reduce *S*'s efficiency.

[1]Confirmation of this conclusion appears in the fact that there was a striking improvement in the school grades of these children after the experiment and that they surpassed a group of undrilled class-mates of their age in performance with the Binet card of objects (Test 32) when tested nearly a year after the period of drill.

(*c*) *Individual capacity* of *S* to attend to, and to assimilate, the particular *type of material* in use—drawings, nonsense syllables, etc.

(*d*) The *ease of assimilation* of the particular test-card in use. Thus, an easy bit of poetry will increase the performance of all *S*'s; a drawing that can be *named*, however fancifully, can on this account be held longer and reproduced better by most *S*'s.

(*e*) *Obstruction or distraction:* Some feature in the object displayed, whether important or trivial, will often catch *S*'s attention, interfere with his exploitation of the balance of the exposure field, and thus measurably reduce his performance. Thus, a misprint in a line of poetry, or the presence of some unusual word, will induce most *S*'s to reread the line, even at the obvious expense of their record.

(*f*) *Ideational type:* Visually-minded *S*'s hold drawings, pictures, and objects by their visual appearance, and are inclined to use visualization for the reproduction of at least portions of the nonsense syllables and poetry. Auditory-minded *S*'s hold verbal material by auditory imagery: if decidedly auditory-minded, like the author, they may also attempt to hold even pictures, drawings, and objects in so far as possible in auditory terms by using verbal formulations, names, etc., as cues for recall. It is important to realize that this transfer from the presented sense-department to some other one more natural to *S* does not necessitate inferior performance. Thus, Foster remarks that while the work of observing visual material "may be surrogated, transferred to a given sense-department," yet "such surrogation may even be advantageous." The best *S* in his tests of visual reproduction was the one who visualized least.

(*g*) *Restriction:* *S*'s performance is definitely conditioned by his voluntary attempt to assimilate and reproduce either a large amount or a small amount of the material exposed. Thus, the nonsense syllables exposed are ten in number: by an effort, all ten may sometimes be read, but the result will be a poor reproduction of two or three syllables; if *S* confines his attention to the first four syllables, he may read these twice, and succeed fairly often in reporting all four. Similarly, there may be voluntary restriction of attention to other types of material.

(*h*) *Grouping:* As noted in Test 24, visual apprehension is greatly facilitated by any device that permits the grouping of the constituent elements in the exposure-field. This factor, more than any other, gives us the explanation both of the individual differences and of the practise-improvement above mentioned. Thus, in the dot tests, the mass of irregularly arranged dots is, by most *S*'s, arbitrarily rearranged (subjectively) into two, three, or sometimes four, groups of dots—each group containing three to six dots. Drawings are, similarly, often split up into component elements, and then recombined by a sort of analysis and synthesis. Sometimes this process is accompanied by the application of verbal symbols as tags for recall, *e.g.*, "a rectangle, two peaks, and an arrow." For most *S*'s, this analytic-grouping method turns out to be more effective than the 'steady stare' which they are prone to employ at first.

REFERENCES

(1) Catherine Aiken, Methods of mind-training. New York, 1896. Pp. 75. For supplementary exercises, see her Exercises in mind-training. New York.

(2) K. M. Dallenbach, The effect of practise upon visual apprehension. To appear in *JEdPs.*

(3) W. S. Foster, The effect of practise upon visualizing and upon the reproduction of visual impressions. *JEdPs*, 2: 1911, 11-22.

(4) Life of Robert Houdin, the king of conjurers, by himself. Phila., 1859. (Ch. 14, p. 256.)
(5) J. P. Hylan, The distribution of attention. *PsR*, 10: 1903, 373-403 and 498-533.
(6) E. E. Jones, Individual differences in school children. *PsCl*, 6: 1913, 241-251.
(7) D. O. Lyon, A rapid and accurate method of scoring nonsense syllables and words. *AmJPs*, 24: 1913, 525-531.
(8) W. B. Pillsbury, Attention, London and N. Y., 1908. Pp. 330.
(9) J. Quantz, Problems in the psychology of reading. *PsMon*, 2: No. 1, 1897.
(10) E. C. Sanford, A course in experimental psychology. Boston, 1895 and 1898. Pp. 449.
(11) G. M. Whipple, The effect of practise upon the range of visual attention and of visual apprehension. *JEdPs*, 1: 1910, 249-262.

TEST 25A

Apprehension of topographic relationship: Spot-pattern test.—It was pointed out in Tests 24 and 25 that visual apprehension is facilitated by grouping and that in the test of counting dots the irregularly arranged dots are by most *S*'s rearranged (subjectively) into groups of dots. If we wish to test more directly this capacity for locating the components of a complex visual stimulus, we may proceed after the method developed by McDougall and used by several English investigators under the name of the 'spot-pattern test.' This test of "scope of apprehension" (McDougall) or "scope of attention" (Burt) or "power of concentration" (Schuster) or "apprehension of topographic relationship," as it has been variously termed, differs from the simpler dot-counting of Test 25 in the following particulars: (1) the dots or 'spots' are limited in number (7 to 10); (2) their number is known to *S*; (3) they are located on the exposure card in such a way as to fall at the intersections of imaginary cross-lines a quarter-inch apart and within an area 1.5 inches square; (4) *S*'s task is to reproduce the position of these spots correctly upon a sheet of quarter-inch cross-section paper (Fig. 61); and (5) the exposure is repeated until this task is correctly performed.

This test has special interest on account of the tolerably high correlations it affords with estimated general intelligence.

Two methods are described, the first a more exact one for use in the laboratory, the second a simpler one for schoolroom use.

A. WITH THE DISC TACHISTOSCOPE

Apparatus.[1]—Disc tachistoscope, with lamp, clamps and head-rest, as in Test 24. Special set of 20 9-cm. square stimulus cards, comprising 20 spot-patterns, 5 each of 7, 8, 9 and 10 spots. Prepared sheets of cross-section paper. Stop-watch.

Preliminaries.—Set up the tachistoscope as directed in Test 24, using the smaller opening in the front screen, but adjusting the instrument to give a point exposure of 2 sec. This may be timed with sufficient accuracy with the aid of the stop-watch, and will be secured approximately by opening the sector to 115 deg. and setting the upper weight at 5 cm., and the lower at 87 cm. With a pencil block off a number of sheets of cross-section paper to outline a series of 1.5 inch squares, as in Fig. 61.

Method.—Seat *S* so as to bring his eyes directly in front of the exposure opening, at some 45 cm. distance and preferably slightly above its level. Give him substantially the following instructions: "About 1.5 sec. after I say 'ready,' there will appear in this opening for a brief period, 2 sec., a card containing 7 [or 8 or 9 or 10] spots. The card will look something like this one [showing another card for a moment]. You are immediately afterward to try to reproduce this arrangement of spots upon this square of cross-section paper [see Fig. 61]. The spots are so placed that each one will fall at the point of intersection of some pair of these lines, that is, at the corner of one of the small squares, but never on the outer margin of the large square. If you fail to get the pattern right at the first trial, the exposure will be repeated until you do."

---

[1]The apparatus and method employed by McDougall, Burt and Schuster are needlessly complex. A special portable tachistoscope was used, in which the exposure cards were illuminated from behind by transmitted light from a lamp shining in a darkened room. The cards were so contrived as to present a uniform 'dead' surface, save when the opening of the tachistoscope shutter illumined them for an instant (0.04 sec.), when the pattern flashed out like a tiny constellation of stars. The exposures were arranged in groups of five, separated by intervals of 1.5 sec., and *S*'s attempt at reproduction was made only after each group of five exposures. (See Burt, 1, pp. 150-151, for further details.) As stated in our results, the rank-order for *S*'s is virtually identical by the simpler procedure we recommend.

Use one card of the simplest patterns, inserted 180 deg. from its regular position, for a short fore-exercise to accustom *S* to the task.

S is always informed of the exact number of spots to be exposed on each stimulus card. A fresh square of cross-section paper is to be used for each attempted reproduction and all preceding attempts are to be kept covered from his sight as soon as they have been filled out.[2] Each trial is numbered in order, and *S*'s capacity is scored directly from the number of exposures required for an absolutely correct reproduction. After each trial *E* will notify *S* either 'right' or 'wrong,' but never anything more than that.

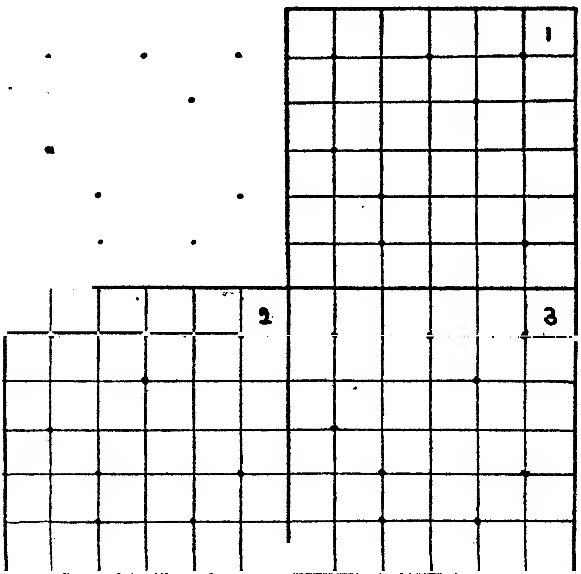


FIG. 61. SPOT-PATTERN AND REPRODUCTIONS.
The original pattern is shown for convenience with the series of three reproductions.

If it is desired to continue the test beyond five patterns of a given number of spots, the standard set of cards may be used

[2]It is not clear whether the English investigators covered *S*'s attempts or not. If not, the conditions under which their results were secured are decidedly different from those here prescribed.

over again three times by turning them about to bring the other edges at the top.

Since the results vary decidedly with the individual patterns as well as with the number of spots, it is important for comparative purposes to give all *S*'s the same patterns in the same position and in the same order. It will depend on the general proficiency of the *S*'s to be examined whether 7, 8, 9 or 10-spot patterns are used. Adults can work with 10-spot patterns after a short period of practise, but 8 and 9-spot patterns are rather difficult for many school children.

B. WITH THE PORTABLE TACHISTOSCOPE

APPARATUS.—Special portable tachistoscope, consisting of a 3-inch, focal-plane photographic shutter, with wire release, mounted on a small easel adapted for holding the stimulus cards. Stimulus cards and cross-section paper as in Method A. Seconds pendulum.[3]

PRELIMINARIES.—Place the shutter on the easel with its curtain at the back, *i.e.*, next the stimulus card. Set the apparatus in such a position as to command full and uniform illumination of the cards when the shutter is open. Place the pendulum in view of *E*, but out of the range of *S*'s vision.

METHOD.—Seating of *S* (save for the head-rest), instructions, fore-exercise, choice of cards, their reproduction and the scoring as in Method A. Wind up the exposure curtain while the face of the apparatus is turned away from *S*: 1.5 sec. after a warning 'ready,' expose the stimulus card and hold the shutter open for 2 sec., using the pendulum for controlling these periods. While *S* is attempting his reproduction, *E* will find it convenient to wind the shutter half-way up, so as to leave the pattern exposed to his view for comparison with *S*'s work. If *S* makes any error, finish winding up the shutter, replace the apparatus before *S* and continue, as in Method A, until a correct reproduction is made.

RESULTS.—Working with trained college students, Miss Goudge obtained the averages given in Table 37 for the disc

[3]As the chief object in devising this apparatus is to secure portability for schoolroom testing, the regular laboratory pendulum may be replaced by a small weight and a cord of the proper length to swing seconds.

tachistoscope (with 1.5 sec. exposures) and for McDougall's method of groups of flashes, with 10 patterns of each type in both methods. As each one of the five flashes in each group is here scored as one exposure (not as 5, as by the English investigators), it will be seen that the 1.5 sec. exposure with the disc tachistoscope is less effective than a group of five exposures of 0.04 sec. each by the McDougall method, but more effective than one of these shorter exposures. Her results also indicate very high correlations between the rankings of the several *S*'s in the two forms of the experiment—0.92, 0.99 and 0.87 for the 8, 9 and 10-spot patterns, respectively. Hence it follows that, at least for adults, the simpler method of single exposures with the tachistoscope furnishes an index of *S*'s ability just as reliable as that given by the more complex method used by McDougall, Burt and Schuster.

TABLE 37.

Mean Number of Exposures Needed to Reproduce Spot-Patterns (*Goudge*)

SUBJECT	DISC TACHISTOSCOPE						M'DOUGALL'S METHOD					
	A	B	C	D	E	F	A	B	C	D	E	F
8 Spots	2.0	4.7	8.5	5.4	7.4	3.3	8.5	12.5	21.0	19.5	22.5	18.5
9 Spots	1.8	2.6	6.5	5.4	3.7	2.6	5.5	9.5	23.0	20.0	16.5	9.5
10 Spots	1.9	2.4	5.1	2.9	3.2	2.7	9.0	9.0	20.5	16.0	13.5	15.0

Burt found 8 and 9-spot patterns too difficult for many 13-year-old pupils. His test with these pupils in an elementary school, when four series of 7-spot, one series of 8-spot and one of 9-spot patterns were used, yielded for boys an average number of exposures (counting each flash) of 45.3, m.v. 21.1, extremes 10 and 157, and with pupils of the same age at a preparatory school an average for 8-spot patterns of 38.6 exposures.[4]

(2) *Dependence on sex.* Burt and Moore found boys superior to girls: median for boys 51 exposures, for girls 64 exposures; percentage of boys exceeding the median performance of girls 65.5, leaving an excess superiority for boys of 15.5 per

[4]It should be pointed out again that it is not clear whether these pupils had their earlier attempts at reproduction in sight as they continued the test.

cent. This test, according to these investigators, constitutes the chief exception to the rule that sex differences are least in tests that correlate highest with intelligence, possibly because the technique proved difficult and the coefficient of reliability correspondingly low.

(3) *Dependence on practise.* Observations of adult *S*'s all show that there is a very rapid improvement due to practise, which usually comes early in the tests. This improvement is chiefly due to the development of schemes of grouping. In Table 37 the fact that Miss Goudge's *S*'s worked first with the 8-spot, then with the 9-spot, and lastly with the 10-spot patterns is a main reason for the result that the 9-spot required fewer exposures than the 8-spot patterns, and the 10-spot less, with one or two exceptions, than the 9-spot patterns.

(4) *Dependence on number of spots.* Taking a number of patterns and averaging the results of several *S*'s, it will be found that the difficulty of a pattern increases with the increase in the number of spots. Thus, tests by Method A in the author's laboratory give as the mean number of exposures for adults: 8 spots, 3.93; 9 spots, 4.96; 10 spots, 6.13: in other words, the difficulty of 8, 9 and 10 spots is practically in the relation 4, 5 and 6. On the other hand, it is possible to arrange a 10-spot pattern that shall be easier than an 8-spot or even a 7-spot pattern, since much depends upon the position of the spots with reference to one another. A card that favors grouping when viewed by an *S* who has developed a satisfactory scheme of mental grouping may be reproduced correctly after a single exposure.

(5) *Correlation with intelligence.* The test has a high correlation with estimated intelligence. By the amalgamation of three series of trials Burt obtained in one school .76, *P.E.* .05, and in another school .75, *P. E.* .09. Again, when the boys in the elementary school were divided into three sections, clever, average and 'infra-normal,' the mean number of exposures was 19.5, 41 and 71.6, respectively, while a weak-minded boy needed 155 on one occasion and 220 on a second. It would seem, therefore, that the spot-pattern test promises to be of diagnostic value.

Schuster's results with a limited number of Oxford students, based on 9-spot patterns by the McDougall method, indicate small, but regularly positive correlations (about .20) between capacity in the test and intellectual status, as determined by a very rough classification according to the 4-fold table method.

(6) *Other relations.* Schuster reports that Oxford students specializing in science and mathematics were superior to those specializing in other subjects, and he also shows that "eyesight does not appreciably affect the results of this test."

(7) *Reliability.* The coefficient of reliability reported by Burt is rather low, about 0.50, but this is explicable by the unnecessarily complex method and apparatus he employed: the high correlations obtained by Miss Goudge between two forms of the test indicate, on the contrary, a very high degree of reliability.

(8) *Qualitative analysis.* The correct reproduction of the spot-patterns necessitates a process of analysis and synthesis. Miss Goudge's introspective records show that the patterns are always learned by a mental grouping of the spots. Usually, *S* gives up the attempt to assimilate the entire pattern as a whole and confines his attention in successive exposures to different subjectively formed groups within the total pattern. Some *S*'s continue for several seconds after the exposure to study the memory after-image of the pattern. The reproduction is assisted not only by visual imagery, but also by kinesthetic factors (movements of eye or hand, felt or imagined) and by verbal formulation. Patterns in which there are one or more rather isolated spots, even if the total number of spots be relatively small, offer special difficulty in grouping.

REFERENCES

(1) C. Burt, Experimental tests of general intelligence. *BrJPs*, 3: 1909, 94-177, especially 150 ff.

(2) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.

(3) Mabel Goudge, A simplified method of conducting McDougall's 'spot-pattern test.' To appear in *JEdPs*.

(4) W. McDougall, Physiological psychology. London, 2d ed., 1908. See especially 129 ff.

(5) E. Schuster, First results from the Oxford anthropometric laboratory. *Biometrika*, 8: 1911, 40-51.

TEST 25B

The form-board.[1]—The form-board seems to have been first used by Seguin as a means of training feeble-minded children, but of recent years it has become better known as a mental test.

The task set by the form-board, that of perceiving or recognizing ten different forms, either by sight or by touch, and making a definite movement of reaction with each form (inserting it into its appropriate 'hole') is such that it is difficult to classify the test precisely. With quite young children and with feeble-minded children the perception of form is doubtless a prime factor, and with older children and adults the same thing is true when the test is carried out blindfolded, but otherwise performance in the test is for them largely conditioned by speed and coördination of movement. Yet again, repetition of the test under proper conditions may make it available as a test of learning capacity.

This complexity in the mental processes concerned in the test is reflected in the statements of those who have made most use of it. Norsworthy, for instance, called it a "test of form perception and rate of movement," and also sought to secure indication of learning capacity from her data. Jones likewise used the test to determine learning capacity, and speaks of it, too, as "a very good test of native ability." This idea that the test has diagnostic value in examining intelligence is again reflected in Norsworthy's statement that "this test seems to me to measure to a certain extent the ability of dealing quickly and well with a new situation" (which approximates Stern's definition of intelligence), and in Witmer's statement that "the form-board is one of the best tests rapidly to distinguish between the feeble-minded and the normal child," to which he adds that "it very quickly gives the experimenter a general idea of the child's powers of recognition, discrimination, memory, and coördination," while "repetition of the experiment leads to a conclusion as to his ability to learn." Wallin believes that the form-board test throws light upon the patient's ability to identify forms visually, upon his constructive capacity and his power of muscular coördination. Goddard says: "We have in our laboratory no other test that shows us so much about a child's condition in so short a time." His table of norms suggests strongly that the test can be of direct service in the examination and classification of mentally defective children.

From what has just been said it is evident that if we wish to measure perception of form and not mere capacity to make rapid movements, the test must be administered to older chil-

[1]The author is indebted to E. A. Doll, Research Assistant at The Training School, Vineland, N. J., for the general arrangement of this test.

dren and adults by a different method than to young or feeble-minded children. Two methods are therefore described, the one for visual, the other for tactual perception. *E* should use the former for children of perhaps 7 years old or less and for feeble-minded children, the latter for adults and for children above 10 or 12 years; for children between 7 and 12, *E* must use his judgment which method to employ. In case of doubt, the tactual should be tried first.

A. VISUAL METHOD

Materials.—Form-board (either Goddard's, Fig. 62, or the new Cornell pattern).[2] Printed record-sheets. Stop-watch.

Preliminaries.—*E* should first learn by heart a scheme of numbering of the blocks and their corresponding holes. The numbering runs from 1 to 10, beginning at the upper left-hand corner (Maltese cross) and proceeding from left to right.

Method.—Let *S* stand before a table of convenient height with the board squarely in front of him and with the blocks placed at the right side for right-handed and at the left side for left-handed *S*'s. The standard position of the blocks for beginning the test with right-handed *S*'s is shown by the following numbering:

	2		4		10
(Board)	7		9		6
	5	8		3	1

For left-handed persons the blocks are reversed to bring them in the same order moving outward from the board. As

[2] The form-board in most common use is Goddard's adaptation of the Norsworthy board. In it the Norsworthy hexagon and octagon are replaced by a Maltese cross and a 5-pointed star. The board used by Witmer is one such as would be obtained by rotating the Goddard board 180 deg. about its lengthwise axis, and is also supplied with a light rim (4, p. 249.)

Mr. D. K. Fraser has pointed out to me that for extended tests with adults by Method B it would be a great advantage to have a board in which the holes were made in a series of removable blocks of such dimensions as to permit of rather free interchange within the board as a whole, which would then constitute a sort of tray, as it were, to contain these ten blocks. Such a device would permit *E* to make various groupings of these ten holes (or of yet others at will) and thus study *S*'s capacity to learn new arrangements. The adjustable board recently perfected in the Cornell Educational Laboratory to meet these conditions is recommended as a useful substitute for Goddard's board.

is indicated by the arrangement of the record sheets, the blocks are picked up beginning with the lower row and working from the board outward, *i.e.,* following the sequence: 5, 8, 3, 1, 7, 9, 6, 2, 4, 10.

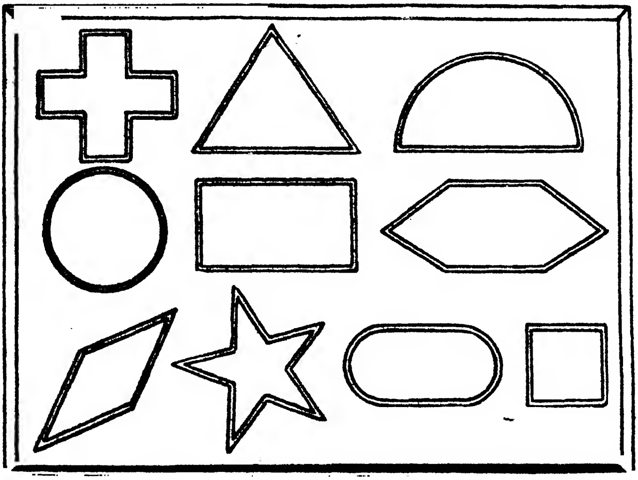


FIG. 62. GODDARD'S FORM-BOARD.

Instruct *S* as follows: "You see these blocks and the holes in the board into which they fit. You are to put the blocks into the holes as fast as you can. You must take the blocks with your right [left] hand, one at a time, in this order from left to right [right to left for left-handed *S*'s], beginning with the lower row [pointing to the blocks to illustrate what is wanted].[3] You must place each block properly before you pick up the next one. Do you understand? Ready. Go!" At this signal *E* starts the watch and records the time taken to place all the blocks correctly.

[3]If *S* is too immature to understand these directions, *E* must push the blocks over to him, one at a time, in the prescribed order. The form-board test has usually been conducted without any attempt to standardize the position of the blocks at the outset, or the order in which they are taken, but experience shows that the outcome may be much affected by chance variations in this respect, so that the regulation order should be followed in every trial.

This quantitative record ought, however, to be supplemented by a qualitative record which shows what precisely *S* does with each block in his attempts to place it correctly. For this purpose *E* uses the printed record-sheets and sets down *S*'s attempts in the following manner. As *S* attempts to place a given block, *E* records against the number of that block on the printed sheet the numbers of the holes into which *S* tries to place it. If a block is correctly placed without any false moves, no number will appear: otherwise, the series of numbers recorded against a given number indicates the full series of trials until the block is rightly placed. The following sample partial record will make the plan clear:

5 — 9 — 3 — 5
8 — 1 — 10 — 1 — 8
3 — 6 — 3
1

Here *S* first tried to place the rectangle in the 'oval,' then in the semicircle, then rightly in the rectangle. Again the third block, the semicircle, was tried in the hexagon before being correctly located. The fourth block, the circle, was placed at once without error, etc.

After this first trial with the 'favored' hand, place the blocks on the other side of the board (left for right-handed *S*'s) in the positions above prescribed (reversed so as to produce the same order moving outward) and repeat the experiment as before, save that *S* must now use the other hand.

A third trial is made with the blocks at the top of the board (arranged in the order of their first trial), and this time *S* is instructed to use both hands.

VARIATIONS OF METHOD.—See below, under Method B.

B. TACTUAL METHOD

MATERIALS.—As before, with the addition of a blindfold for *S*.[4] Large cloth to cover board and blocks.

METHOD.—Keep the material covered with the cloth when *S* is not working with it and on no account permit him to see

[4]The blindfold must positively exclude any possibility of glancing at the board from underneath its folds. For children *E* will find it desirable to construct a special blindfold that will guarantee perfect exclusion of the sight of the board.

the board or the blocks before or during the test, or after it if there is any likelihood of later trials. The blocks are arranged as in the visual method, but off to one side, under *E*'s charge, who shoves them over to *S*, one at a time, in the order prescribed.

When *S* has been properly blindfolded and led to the position in front of the board, instruct him thus: "On the table before you there is a board in which are a series of ten holes of different shapes. I shall hand you, one at a time, a series of blocks of wood that just fit these ten holes. Put each block in place as quickly as you can. A new block will be placed for you just here [indicating a point on the table just at the right of the board] as soon as you have placed the one before it."

S is allowed to handle the block and examine the board with his fingers in any way that he wishes. Note whether both hands are used, and record any points of interest in *S*'s way of accomplishing his task, particularly whether he continues to seek 'blindly' to thrust the block into some hole or uses his hands to feel the holes first. Keep both the quantitative and the qualitative records as in Method A. In the qualitative record, exploring the holes with the fingers does not, of course, constitute an error, but only the attempt to insert a block into a wrong hole.

When the last block is correctly placed, cover the board carefully with the cloth, remove *S*'s blindfold and let him try to sketch the positions of the forms and their shapes.

Restore the blindfold and repeat the test precisely as before. Make a third similar trial or continue the test in this manner until *S* is able correctly to reproduce all the forms in their proper positions. Note the number of trials made and plot a curve of time taken in relation to number of trial.

VARIATIONS FOR BOTH METHODS.—(1) In either the visual or the tactual method the board may, at any point in the series of trials and without previous warning, be turned about 90 or 180 deg., and the effect of this alteration noted. (Cf. the special 'adaptation-board' devised by Goddard for a simpler, but analogous test of low-grade intelligence.)

(2) Some indication of memory type may be obtained by permitting *S* to study the board visually (without inserting the blocks) until he has learned all the forms and their positions (as determined by a sketch made from memory). Then blindfold him, turn the board through 90 deg., and require him to insert the blocks after having been notified of the alteration made in the position of the board. Note whether he seems to have visual or kinesthetic memory, whether he makes errors, whether he is forced to use the method of trial and error, etc.

(3) In Method A the number of trials may be increased to 30, 10 for each hand and 10 for both hands, in regular alternation of 'favored' hand, other hand and both hands. These trials may then be separately averaged and the results taken as measures not only of learning capacity, but also of dexterity, though for the latter purpose the first few trials, while practise is being acquired, must be excluded.

TREATMENT OF DATA.—Unless it is evident that there has been some disturbing factor that should have been eliminated, *S*'s quantitative performance in the first trial may be taken as the measure of his normal unpractised performance.[5]

The qualitative record, in terms of numbers of errors (each attempt to place a block in a hole in which it does not belong constitutes one error) may usually be taken as the measure of his inability to perceive or recognize form. In Method A persistent attempts to insert a block where it is manifestly impossible for it to go, or such absurd things as turning the blocks upside down to make them fit, standing them on end, etc., should be especially noted, as they are symptomatic of decided immaturity and are often seen in mentally defective *S*'s.

RESULTS.—(1) *Norms for Method A.* Table 38 shows the results published by Goddard from a study of 271 normal children classified by chronological and by mental age and of 420 mentally defective children classified by mental age. The same results are shown graphically in Fig. 63. They are based upon Method A, speed in seconds, best of three trials, without distinction of sex. They are probably subject to some revision

[5]For direct comparison with Goddard's table and curve, *E* must, however, take the best of the first three trials.

TABLE 88

Mean Times for the Form-Board, Visual Method (*Goddard*)

NORMAL			NORMAL			MENTALLY DEFECTIVE		
Chron. Age	No. Tested	Mean in Sec.	Mental Age	No. Tested	Mean in Sec.	Mental Age	No. Tested	Mean in Sec.
--	--	----	4	7	33.8	4	53	76.12
5	17	29.5	5	7	30.3	5	52	51.25
6	26	27.5	6	13	27.5	6	54	38.24
7	25	24.5	7	47	25.4	7	85	26.39
8	28	21.8	8	43	20.7	8	87	23.80
9	47	19.3	9	46	19.2	9	48	18.30
10	49	18.2	10	69	16.6	10	29	17.50
11	38	17.6	11	25	15.9	11	8	16.40
12	20	15.9	12	14	14.3	12	4	12.00

and unfortunately furnish no clue to the amount of variation about the means for each age. From supplementary data obtained from Vineland it would appear that the mean variations range roughly from 2 to 5 sec. in the years 8 to 12 and approximate 40 or 50 sec. (for the defectives) in the mental ages 3 to 5. Children whose mental age is 1 or 2 years almost invariably fail with the test, and a certain percentage of those mentally 3 and 4 years also fail entirely. Attention should be called to the fact that in ages 4, 5 and 6 there is a marked discrepancy between the speed of normal and of feeble-minded children of the same mental age.

For data on the form-board test as applied to epileptics the elaborate tables prepared by Wallin, showing relation to age, sex and types of feeble-mindedness, should be consulted.[6]

(2) *Norms for Method B.* Data secured by the author from a small number of advanced university students give for three successive trials without sight of the board averages of 140, 120 and 69 sec., respectively. When the attempt was made to sketch the board after each trial, it was found that no *S* could accomplish this after the first trial, whereas about one-half located the blocks correctly and sketched their forms with but

[6]Wallin's averages are not entirely reliable in the totals, because some cases have been omitted entirely and none of these totals are weighted averages—a fact which would perhaps explain certain contradictions which appear in them.

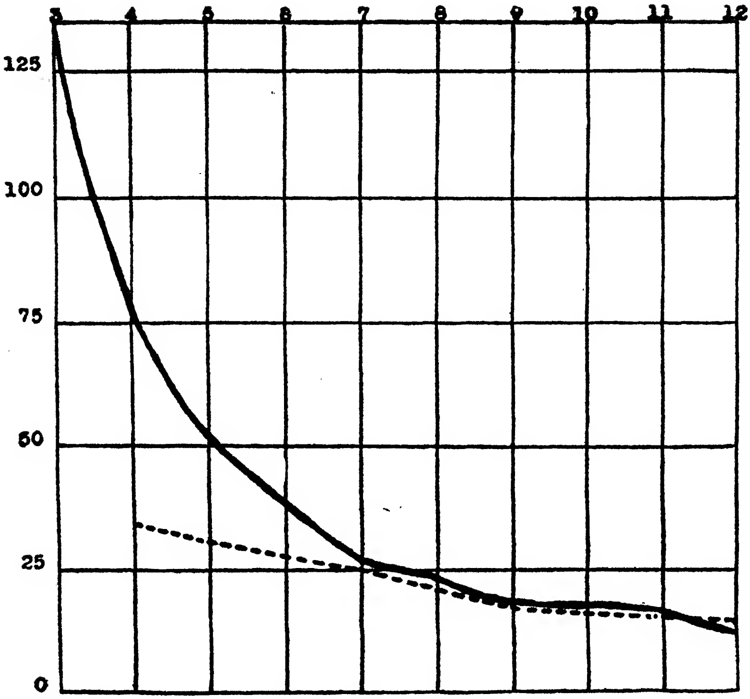


FIG. 63. MEAN TIMES FOR THE FORM-BOARD (From Goddard).
Ordinates represent seconds, abscissas mental age, solid line mental defectives, dotted line normal children.

trivial errors after the third trial. Decided individual differences are discovered in capacity to identify forms tactually. It is possible for *S* to learn to place the blocks rapidly by touch and yet have extremely erroneous ideas of their actual form.

(3) *Dependence on practise.* Some idea of the improvement due to repetition of the test (visual method) may be gained from Jones' figures for 15 unselected 4th-grade pupils, from which the following figures may be obtained:

	average	m. v.
1st trial,	22.3 sec.	4.4 sec.
3d trial,	19.3	3.3
10th trial,	15.5	2.7
Mean,	18.2	2.4
Best,	14.2	1.8

Jones found that the best record was made in 5 cases on the 10th trial, in 5 cases on the 9th, in 3 on the 7th, in one of the 4th, while in one both the 5th and the 10th were equally good. The maximal improvement for any pupil was a reduction from 32.4 to 10.6 sec., the minimal from 18.2 to 17.2 sec.

Norsworthy's repetition of the test after a year's interval showed the median improvement to be 1.2 sec. for defective, and 10.0 sec. for normal children.

REFERENCES

(1) H. H. Goddard, The form-board as a measure of intellectual development in children. Reprint from *The Training School*, 9: June, 1912. Pp. 4. (See also The adaptation-board. *PsBu*, 9: 1912, 79-80.)

(2) E. E. Jones, Individual differences in school children. *PsCl*, 6: 1913, 241-251.

(3) N. Norsworthy, Psychology of mentally deficient children. *ColumbiaConPhPs*, 1906. Pp. 111.

(3*a*) R. H. Sylvester, The form-board test. *PsMon*, 15: 1913, No. 65. Pp. 56.

(4) J. E. Wallin, Experimental studies of mental defectives. *EdPsMon*, No. 7, Baltimore, 1912. Pp. 155.

(5) J. E. Wallin, Human efficiency. *PdSe*, 18: 1911, 74-84.

(6) L. Witmer, A monkey with a mind. *PsCl*, 3: 1909, 179-205.

(7) L. Witmer, Courses in psychology at the summer school of the University of Pennsylvania. *PsCl*, 4: 1911, 245-273.

TEST 26

Cancellation.—There are in use several forms of mental test in which a continuous task is assigned under conditions such that maximal attention is demanded for the best work, and that any reduction of attention is reflected directly in the speed or accuracy of the work. Prominent among these tests is that which, following Miss Sharp (27), we shall term 'cancellation.' The essential principle is the crossing out of an assigned symbol or symbols (letters, words, digits, etc.) from a prepared form containing the assigned symbol in conjunction with a number of others of the same general order. The test is rather remarkable for the variety of forms it has assumed, the variety of names that have been given it, and the divergence of statement as to what it really measures.

Thus, its originator, Bourdon (6), used it with ten adults to measure "discrimination;" Oehrn (24) proposed it, under the title "search for assigned letters," as a convenient test of attention for experimentation in individual psychology; Cattell and Farrand (10) introduced it into the series of tests of Columbia University students in the form of the "*A*-test" for "rate of perception," and Thorndike employed it later, together with what may be regarded as variations of it (the "*a-t* test," the "*e-r* test," and the "misspelled-word test") for various comparative studies, particularly for his examination of the mental traits of twins

it to contrast the ability of intelligent and unintelligent children (as did Winteler, 37), to measure fatigue (as did Ritter, 25), and incidentally, to study their capacity to break and form associations; Vogt used cancellation to examine attention and distraction; Judd (17) has classed it among tests devoted to discrimination reactions, Woodworth and Wells as an association test; Simpson speaks of it as a test of "efficiency of perception," Chambers as a test of "quickness of perception," Wallin as a test of "speed and accuracy of perceptual discrimination," Franz as a test of "time of discrimination, association and movement," while Hollingworth points out that the processes depend somewhat on the subject: for some there is mere recognition of the assigned symbol without much discrimination, whereas for others there occurs deliberate discrimination of that symbol from the other symbols. Meumann objects to all these general characterizations and insists that we must discover the elementary processes at work. The primary activity demanded of the subject in the cancellation test is, he says, the holding in attention of the problem assigned him, and this means that he must keep in mind not only the particular letters to be cancelled, but also what is to be done with them (noting their presence and crossing them out). 'Discrimination' or 'quickness of apprehension' is something secondary to this primary activity. But this keeping in mind of a specific thing or attribute to be looked for in a totality of presented material is, he says, the fundamental activity in all systematic observation, so that the test can be said, if we would use a general term, to measure real capacity of observation.

In the opinion of Bobertag[1] the value of the test is not very clear, since it certainly does not test pure attention, and seems particularly affected by non-intellectual factors, such as conscientiousness and good-will.

However the activity concerned in cancellation may be described in psychological terms, it is clear that the objective results of the test must vary, even for the same *S*, according as the test-conditions are varied with respect to (*a*) kind and arrangement of symbols, (*b*) number of symbols assigned to be cancelled and (*c*) the duration of the test. To secure direct comparison with the work of others for material and method other than that prescribed below, *E* must consult the original sources for the detailed description of test-conditions, but some idea of the methodological variations that have been used may be given at this point.

(*a*) *Kind and arrangement of symbols.* We may distinguish at least five kinds of material with variations in the arrangement of some of these kinds: (1) spaced prose, (2) unspaced prose, (3) disconnected words, (4) unregulated pied material, and (5) regulated pied material, with (*a*) unequal, or (*b*) equal, proportion of the symbols.

(1) *Spaced prose* is secured by taking any ordinary printed matter, but preference has been given to scientific (Binet, Lapie) or philosophical (Sharp) texts, rather than to easily grasped material (Winteler).

(2) *Unspaced prose*, as employed by Miss Sharp, is secured by having a page of concrete description, or of more abstract material, printed without capitals, spacing, or punctuation. The following lines will serve to illustrate two such texts:

---

[1] Expressed in a personal communication to the author.

theshoresoftheislandwerecompletelyfringedwithbushesand
greatcarehadbeentakentopreservethemastheyansweredasascreen

theproblemofphilosophyhasbeenineveryagetodeterminethe
relationbetweenbeingandthoughtbetweensubjectandobjecteveryphil

(3) *Disconnected words* may take the form of (*a*) sense words in nonsense arrangement (Sharp) or of (*b*) nonsense words, as in the a-test used by MacMillan and Bruner and by Mrs. Squire. Similarly, the arrangement of French words in irregular order as used by Brown with English children amounts to nonsense arrangement as does the style of form illustrated in the misspelled-words test prescribed below.

(4) *Unregulated pied material* has sometimes been formed, after the fashion of Bourdon and Vogt, by printing a page of some foreign language sure not to be known (Hungarian, Finnish) without spacing, capitals or punctuation, but the same material can be secured more simply by letting the printer set up a page of letters picked up at random from a pile of 'pied' type.

(5) *Regulated pied material, i. e.*, an arrangement of symbols in irregular order, but with precaution to secure a given number of certain of them has perfectly obvious advantages for experimental work, and this is the form of material now used by nearly all experimenters. Aside from the choice of the nature of the symbols (letters, digits, etc.) this form of material subdivides into two forms, according to the arrangement into (*a*) unequal and (*b*) equal number of each symbol.

A good example of unequal distribution is seen in the familiar A-test (reproduced herewith, and employed by Whitley, Doll, Simpson, Chambers and as one of the stock tests of Columbia freshmen), in which there are 50 A's interspersed among other letters of less frequency, and other examples in the six forms of 200 A's each used by Wallin, the B-test and geometrical form test as used by Simpson, Wyatt's form with 600 letters, Whitley's form containing 50 each of A, B, K and S (36, p. 62) and her set of hieroglyphic characters (p. 69) and in the curious rows of dots arranged in groups of 3, 4 or 5 dots, 24 groups to the line, used by Abelson.

Equal distribution of the symbols is exemplified in the number-checking tests devised by Woodworth and Wells, in an improved geometrical-form sheet described by Whitley (p. 70) and in Franz's special set of seven forms (15, pp. 130-133), each containing 25 of eight different symbols (letters, digits or other forms) so placed that a given character in the one form corresponds in location with a given character in every other form, which permits special examination to be made of ability to learn (see Kent and Boring).

THE A-TEST

OYKFIUDBHTAGDAACDIXAMRPAGQZTAACVAOWLYX
WABBTHJJANEEFAAMEAACBSVSKALLPHANRNPKAZF
YRQAQEAXJUDFOIMWZSAUCGVAOABMAYDYAAZJDAL
JACINEVBGAOFHARPVEJCTQZAPJLEIQWNAHRBUIAS
SNZMWAAAWHACAXHXQAXTDPUTYGSKGVKVIGKIM
FUOFAAKYFGTMBLYZIJAAVAUAACXDTVDACJSIUFMO
TXWAMQEAKHAOPXZWCAIRBRZNSOQAQLMDGUSGB
AKNAAPLPAAAHYOAEKLNVFARJAEHNPWIBAYAQRK
UPDSHAAQGGHTAMZAQGMTPNURKNXIJEOWYCREJD
UOLJCOAKSZAUAFERFAWAFZAWXBAAABHAMBATAD
KVSTBNAPLILAOXYSJUOVYIVPAAPSDNLKRQAAOJLE
GAAQYEMPAZNTIBXGAIMRUSAWZAZWXAMXBDXAJZ
ECNABAHGDVSVFTCLAYKUKCWAFRWHTQYAFAAAOH

The *A*-test has three disadvantages: it is not devised to permit the use of any letters other than *A;* it does not lend itself easily to four-letter cancellation; in many instances, two or more *A*'s are in juxtaposition, so that some *S*'s may see and cancel several at once.

To avoid these difficulties, recourse may be had to the second form of regulation, viz: the use of each of the 26 letters the same number of times each. In the two letter-forms which are prepared for this test, the printer has mixed together 100 types of each letter and set them in chance order, save where a given letter was repeated too closely.

(*b*) *Letters cancelled.* It is easily demonstrated that the ranks of a given number of *S*'s will vary somewhat according as a few (1-2) or several letters (4-6) are to be cancelled. Where but one letter has been used, this has commonly been the letter *a* (Bourdon, Cattell and Farrand, Wissler, Thorndike, Sharp, Descoeudres, Simpson, Squire, Chambers, Doll, Wallin, Whitley, MacMillan and Bruner) or occasionally *e*. In cancelling more than one letter, we find that Bourdon tried at different times, *a* and *i; a, r, i, s; a, e, l, t;* or *a, e, l, t, o, k;* Binet, *a, e, d, r, s,* also *i, o, l, f, t,* and *a, e, r, o, s, m;* Vogt, *l, n, s;* Winteler, *n, s, t,* and *l, m, r, s, t;* Burt, *o* and *c;* Lapie, *a* and *r;* Wyatt and Brown, *e* and *r,* and *a, n, o, s;* Whitley, *a, b, k, s;* Woodworth and Wells and Strong, any single digit; Kent and Boring, various single characters; and Whitley and Simpson different single geometrical forms.

It was Binet's contention that, when but one letter is cancelled, *S*'s tend to work with approximately equal accuracy, but with varying speed, whereas, when four or five letters are cancelled, they tend to work with approximately equal speed, but with varying accuracy; he, accordingly, arranged the test in one or the other of these ways, as he wished to measure either speed or accuracy. As will be shown later, this assumption is not strictly justifiable.

The cancellation of one letter is so easy that some investigators have tried other devices for complicating the task: Ritter, for example, had his *S*'s cancel every *r* with a vertical stroke and every article (in German texts) with a horizontal stroke, or, again, every *s* with a vertical and every preposition with a horizontal stroke: his idea that these two assignments would be equally easy and could thus be used interchangeably was not borne out in practise. Thorndike has used still another variation: mark every *word* (in a Spanish text) containing both *e* and *r,* or both *a* and *t.* Similarly, in the "number-group checking test" of Woodworth and Wells, *S* is to cross out every combination containing both 2 and 3, or both 8 and 9. Investigators at Columbia University have also made frequent use of the misspelled-word test described below.

(*c*) *Duration.* It is preferable, whenever possible, to have all *S*'s cover the same amount of material and to secure the quantitative record in terms of elapsed time (Abelson, Boring, Kent, Descoeudres, MacMillan and Bruner, Simpson, Squire, Strong, Woodworth and Wells and others). But for group work it is necessary to use the time-limit method (unless using the special time-clock), and the actual limits used by investigators have ranged from 45 sec. to several hours. At Columbia, 1 min. is allowed for the *A*-test, 2 min. for the *a-t* test, 3 min. for the misspelled word test, When working by the individual method, the *A*-test takes, on the average, about 95 sec. Miss Sharp's tests demanded from 3 to 4 min.; Bourdon used 6 min.; Winteler 15 min.; Binet found that by extending the time to 10 or 20 min., he could get signs of fatigue. When such long times are used, *E* may follow the time with the stop-watch and at expiration of each minute, check with a small horizontal underlining mark, the point at which *S* is then working. It was Oehrn's idea that the cancellation test might be extended through several hours in

order to analyze such factors as praise, fatigue, ennui, spurt, warming-up, etc. This has been partially done by Vogt, who worked for an hour or more at a time, in 10 min. periods with 5 min. intermissions.

Among recent investigators we find durations for this test of 75 sec. (Chambers), 100 sec. (Wallin), 2 min. (Doll, Burt, Wyatt), 3 min. (Wyatt's *anos*-test, Brown with adults), and 5 min. (Lapie and Brown with children).

Seven forms of cancellation test are here described: the first four furnish material for the cancellation of one or more symbols, whether letters, digits or geometrical forms; the fifth reproduces the word-cancelling tests and the last two the misspelled-word tests, as administered at Columbia University.

A. CANCELLATION OF A SINGLE SYMBOL

MATERIALS.—Four printed forms. Set of three control keys. Stop-watch. Moderately soft pencil.

The first form, beginning *hplg*, etc., contains 50 each of the 26 letters, arranged in chance order, but with precaution against the repetition of any letter within two places.

The second form, beginning *zycu*, etc., is a substitute form of similar construction.

The third form, geometrical figures, contains 8 different characters, each of which is presented 50 times, and is identical with the form used by Whitley (36, p. 70).

The fourth form, digits, resembles the 'number-checking test' of Woodworth and Wells, but has been improved in several respects. The digits are printed in 12-point instead of 8-point type, with about doubled spacing and leading, and the location of the several digits has been so arranged that by means of but three control keys the location of all 100 of any one of the digits can be rendered instantly visible, thus greatly facilitating the correction of *S*'s performance. Due care has been taken to avoid 'runs' of any digit, so that any digits may be assigned for cancelling.[2]

METHOD.—Whenever convenient, work with one *S* at a time. Place the form before *S*, printed side down. For the first form instruct him as follows: "When I say 'ready,' turn over this sheet of paper, begin at the first line, and mark every *a*[3] on the page like this." (Exhibit for an instant another sample form, already marked, to be sure that *S* understands the instructions.) "Mark as rapidly as you can, but try not to leave out

[2]For the construction of this form and its ingenious keys I am indebted to my assistant, Mr. D. K. Fraser.

[3]The letter *a* is chosen merely because it has commonly been used. If it is desired to repeat the test with other letters, it would doubtless be preferable to select a number which have been shown to be of nearly equal legibility. For this, one may recommend *m* and *w*, or the four letters, *q*, *p*, *b*, *d*. (See Sanford, 26).

any *a*'s." Give the command—"ready." Start the watch when *S* glances at the first line: stop it when he finishes the last line; record on the form the time, together with *S*'s name, the date, and other needed items.

Repeat with another letter, as *o*, on the same form, or with the letter *a* again on the second form.

Test the third form, using similar instructions and assigning the triangle for cancellation. For further testing, the circle or semi-circle will be found easier than the trapezoid.

In using the fourth form, try for the standard test the digit 7, with the digit 1 for the second test. For further testing, 0 and 4 will be found harder than 1 or 7, while 6 and 9 are the hardest to cancel.

S may be allowed to traverse alternate lines from right to left.

If necessary to follow the group method, as in conducting classroom tests, it is a good plan to write a sample line on the blackboard and show how the cancellation is to be done. Make sure that no *S* turns the paper before the signal, and that all start simultaneously. All *S*'s must cease work at the command "stop," and underscore horizontally the letter at which they were looking when this command was given. The exact time to be used must be determined by a few preliminary experiments; it should be such that the fastest *S* to be examined will not quite be able to finish: 2 min. is to be generally recommended, since very few adults can finish within that time.[4]

B. Cancellation of more than one symbol

Materials.—As in Method A.

Preliminaries.—The characters to be cancelled must first be typewritten or drawn upon each form, so as to be readily visible to *S* during the work. For letters it is recommended that *q, r, s, t* be cancelled upon either the first or second form;[5] for the

[4] The author's time-clock for group testing will render group testing by the work-limit method feasible with reliable *S*'s.

[5] Save for special purposes, the cancellation of four letters suffices as well as that of five or six to bring out the characteristics of the test. The letters *q, r, s, t* are selected because they form a combination fairly easily remembered, and embrace one letter projecting above, one projecting below the line, and two small letters.

third form try any two, three or four of the geometrical figures; for digits, try two or more of the digits 2, 4, 6, 8.

Method.—As before, with the added explanation that the letters to be marked have been placed for reference at the top of the sheet. (*E* must, of course, also state the letters verbally beforehand.) The time will naturally be longer than before: 4 min. may safely be used in most group tests, as a competent adult takes nearly 5 min.

Variation of Method.—Omit the typewriting of the letters to be cancelled.

C. Cancellation of words—the a-t and the e-r test

Materials.—Printed page from a Spanish text. Stop-watch.

Method.—Instruct *S* to mark with a horizontal stroke each *word* containing *both a* and *t*. Exhibit a sample page, or illustrate on the blackboard. Forewarn *S* that the words are in a foreign language. If the individual method can be followed, take the time for the whole sheet; if the group method is followed, allow 2 min.

For a second test, use fresh sheets of the same text, but substitute *e* and *r* for *a* and *t*.

D. Cancellation of misspelled words

Materials.—Two printed texts containing a large number of misspelled words. Stop-watch.

Method.—Test each text separately. In each case, instruct *S* to mark with a horizontal stroke every word that is not spelled correctly. Take the time for the whole sheet, or allow 3 min.

Treatment of Data: Scoring.—In these cancellation tests we meet the difficulty common to many mental tests that we have two indexes of efficiency—the one quantitative (speed), the other qualitative (accuracy). There are evidently three ways of scoring under these conditions: (1) we may keep both records and try to use them independently, in which event we may find it almost impossible to make any extensive comparison of results for different individuals or different groups; (2) we may neglect either speed or accuracy and make all our com-

parisons in terms of the other index; or (3) we may attempt by some corrective formula to combine the two indexes into a single index that shall express fairly, though perhaps somewhat arbitrarily, *S*'s net efficiency. Examples of these three methods will make them clearer.

(1) In Doll's tabular presentation of the work of feeble-minded children the average number of *a*'s crossed is given for each group as a measure of speed and the average number of *a*'s passed over (omissions within the portion of the form attempted) is likewise given for each group as a measure of inaccuracy.

(2) Under some circumstances it will be found that the errors are so few as to become virtually negligible, in which case comparison may be made directly in terms of speed (see, for example, the recommendations made by Woodworth and Wells for their number-checking tests). Similarly, MacMillan and Bruner omit their corrective formula if the accuracy is 96 per cent. or above. Binet, as noted above, also believed that the cancellation of one letter tended to produce approximately equal accuracy and that of four or five letters approximately equal speed.

(3) The conversion of speed and accuracy into a single index of efficiency has been attempted in various ways. The simplest is doubtless to select some arbitrary quantity as a penalty for each error and to use this same penalty for all *S*'s. Thus, Simpson corrected the time for the A-test by adding 5 sec. for each *A* omitted, and his geometrical form test by adding 3 or 6 sec. (depending on the assigned symbol) for each omission and neglecting the few errors of wrong cancellation. The principle used is that of penalizing *S* for each omission by an amount slightly greater than the time found by experience to be needed on the average for cancelling one symbol. Another very simple correction is that used by Wyatt and by Brown when working by the time-limit method: give + 1 to each symbol correctly marked and — 1 for each error of omission or of insertion within the ground covered. A similar method is used by Wallin for scoring his forms containing 200 *A*'s: quantitative efficiency is measured by crediting 0.5 per cent. for each *A* cancelled, while combined quantitative and qualitative efficiency is measured by subtracting from the quantitative score 0.5 per cent. for each *A* missed or wrongly crossed within the ground covered.

Other investigators have used more elaborate corrective formulas. Thus Miss Sharp converted speed into accuracy on the assumption that "in a given individual maintaining a constant degree of attention while doing a piece of work the percentage of error is inversely proportional to the time taken for the work." But most investigators, following the suggestion of Cattell and Farrand, have sought to convert accuracy into speed by some formula that would add to the obtained time an amount that would be required to mark the letters omitted, with possibly some additional penalty. The first method may be illustrated by reference to MacMillan and Bruner, who divided the recorded time by the proportion of *a*'s crossed out (unless this fraction was 96 per cent. or over), the second method by the scheme of Strong, who for each omission added to the obtained time twice the time required to mark a single letter.

The following treatment has been used by the writer. When possible, rank speed in terms of time (T), otherwise in terms of symbols examined (e), *i. e.*, ground covered.

Let $A =$ the index of accuracy,
$E =$ the index of net efficiency,
$T =$ the time in seconds for the entire form,
$e =$ the number of symbols examined,
$o =$ the number of letters erroneously omitted,
$c =$ the number of letters crossed,
$w =$ the number of letters wrongly crossed.

Then, to compute the index of accuracy, use the formula

$$A = \frac{c - w}{c + o}.$$

To compute the index of net efficiency when using the time-limit method, employ the formula

$$E = eA,$$

and when using the work-limit method, the formula

$$E = 100 \frac{A}{T}.$$

The chief defect of these formulas seems to lie in over-severe penalizing of extreme cases. Thus, an *S* in the *qrst*-test examined 825 letters, crossed 40 (including 2 *w*'s) and omitted 78. Here $e = 825$, $A = .322$, $E = 266$.

While the author has not had opportunity to test it, the plan proposed by Münsterberg to meet a similar difficulty in another connection would appear likely to supply a still better correction (22, 77 ff.) and possibly to provide an adequate method of equating speed and accuracy in many tests. The idea is to ascertain the approximate range of both quantitative and qualitative performance and to equate units on the qualitative range with units on the quantitative range. To borrow the figures from Münsterberg, assume that in the cancellation test times ranged from 180 to 420 sec., and errors from 4 to 28. Then one of the 24 qualitative units is equal to 10 of the quantitative units, so that a penalty of 10 sec. may be added to the time record for each error in cancellation.

It is often of interest to keep records of the number of omissions for each letter, as in this way one may get comparative estimates of their difficulty.

For the word-cancellation, similar methods may be pursued: when the test is administered by the time-limit method, efficiency, for the sake of simplicity, may be taken in terms of the number of words marked, with a deduction of 2 for each wrongly marked word.[6]

The rank in the misspelled word test may, similarly, be best computed in terms of time, with a reduction for errors as described for the cancellation of letters.[7]

RESULTS. (1) *Average performances.* The number-checking test of Woodworth and Wells takes for adults from 100 to 200 sec., with an average of about 133 sec. The *A*-test at Columbia has yielded averages of 87.3 sec. for women and 100 sec. for men. A similar test, used by Chambers with a time-limit of 75 sec., was scored in *A*'s crossed by 7th and 8th-grade pupils, average 60.25, minimum 39, maximum 95. The misspelled word tests are of unequal difficulty: on the average the time required by adults per word cancelled ranges from 1.11 to 1.87 sec., with a mean (based on 183 students) of 1.64 sec. For further norms consult the tables that follow and those published by Whitley (p. 72).

(2) *Dependence on age.* There is undoubtedly a general improvement with age both in speed and accuracy of cancellation. Wissler's results for the *A*-test show that Columbia seniors better their records as freshmen: in 35 cases there was an average reduction in time from 105.4 to 88.9 sec., and in errors from 4.7 to 1.6. The relation of performance in this test to mental age is shown in the Vineland results published by Doll, and reproduced here in Table 39. It will be seen that defectives classed as 1 to 3 years old mentally cannot perform the test and that total failure may extend up through the 6th year. From the 4th to the 12th mental year the number of *a*'s crossed increases from an average of 11 to an average of 41 (2 min. time-limit).

[6]This method is followed by Thorndike, who found such errors only in one of ten papers on the average.

[7]Thorndike neglected errors in this test, though they occurred in about one-third of the papers.

TABLE 39

Results with the A-test at Vineland, Classed by Mental Age (*Doll*)

MENTAL AGE	AVER. CHRON. AGE	NO. OF CASES	NO. OF TESTS	NO. OF FAILURES	NO. A'S CROSSED		AV. NO. LETTERS OMITTED	EXTREMES	
					AV.	M. V.		WORST	BEST
1	16	7	7	7					
2	17	6	7	7					
3	17	13	17	17					
4	17	22	37	16	11	3	15	0	16
5	22	24	26	12	13	5	20	0	27
6	20	38	61	6	21	8	14	0	51
7	19	32	64	0	18	6	17	4	34
8	20	55	109	0	25	7	12	3	64
9	20	33	89	0	33	9	9	15	50
10	21	14	34	0	35	8	6	17	51
11	19	6	19	0	41	15	4	30	56
12	18	2	9	0	41	4	4	37	45

Mrs. Squire, using the special *a*-test devised by MacMillan and Bruner, found a fairly steady decrease of 'perception-time' with age, with two distinct points of sudden improvement, viz: at the years 7-8 and 11-12. She therefore proposes to include this test with others in a scheme of diagnostic tests, with the following standards of performance.[8] Age 6, omissions (taken to indicate degree of attention) not to exceed 11, time not to exceed 175 sec.; age 7, omissions under 11, time under 90 sec.; ages 8 to 10, omissions under 6, time under 60 sec.; age 11, omissions under 5, time under 60 sec.; ages 12 or 13, omissions not over 1 and time 49 to 50 sec.

(3) *Dependence on sex.* All investigators concur in reporting a sex difference in favor of women and girls. The results of freshmen tests at Columbia have already been quoted. Woodworth and Wells find women distinctly better in the number-checking test; Burt, with 2 min. cancellation of *o*'s and *e*'s, found that boys cancelled 99.5, girls 110.0 letters, while Doll's figures for Vineland show, for the ages where sufficient cases of both sexes are included, a superiority for the girls of from 8 to 16 per cent.

[8]There is at least serious doubt as to the validity of the 'perception-time' here used. It is computed as by MacMillan and Bruner, whose method is criticized below (Result 8).

(4) *Dependence on letter cancelled.* According to Bourdon, letters whose form is simplest are oftenest omitted; with the letter *o*, 6 to 10 times as many errors were made by his *S*'s as with the letter *e*; but the author's comparative tests with the letters *c* and *o* do not substantiate these conclusions fully. Table 40 shows that in 40 consecutive tests, grouped by tens, *o* is distinctly easier to cancel than *c*, save in the first group; it may therefore be expected that it is easier to cancel than *e*, which, in practise, is often confused with *c*. In the *qrst*-test, 50 boys committed errors of omission as follows: *s*, 337; *t*, 561; *r*, 653; and *q*, 718. Here there is no relation whatsoever with legibility, since *s* belongs to the group of relatively illegible letters, and *q* to the group of very legible letters (Sanford, 26). We may conclude that when school children cancel these four letters in one test, it is the least often used letter that suffers most from omission.

TABLE 40

Effects of Letters and of Fatigue on Cancellation (Whipple)

LETTER	1ST TEN	M. V.	2ND TEN	M. V.	3RD TEN	M. V.	4TH TEN	M. V.	AVERAGE
c	93.12	5.36	91.04	5.73	89.86	3.0	96.64	4.69	92.66
o	94.06	4.93	86.28	1.42	82.64	3.4	87.52	3.46	87.62
Average	93.59		88.21		86.25		92.08		

In number-checking Woodworth and Wells find 1 and 7 easiest, next 0 and 4, and 6 and 9 hardest. In crossing *A*, *B*, *K* and *S* in a page of capitals, Whitley found *K* harder than *B* and *A* harder than *S*. Her monograph may be consulted for further details on the subject of this and of the following paragraph.

(5) *Dependence on the form of material.* Bourdon found that a change in size of type from one in which the small letters were 1.75 mm. high to one in which they were 1.25 high made little difference: if anything, the smaller type proved better. On the other hand, several investigators have complained against the use of small type (Strong, for instance, finds even

Woodworth and Wells number-checking form too trying on the eyes), and Whitley found the use of hieroglyphics resembling Chinese letters quite unsuited for cancellation work.

There seems no doubt that more errors are made in cancelling letters out of ordinary prose than out of 'pied' material. Sharp, for example, found that even the elimination of spacing between words altered the performance: the average time for 7 *S*'s for a spaced text was 190 sec., for an unspaced text 219 sec., and the average percentage of errors was 9.6 and 4.58, respectively. The faster speed and poorer quality of work done with ordinary prose is plainly due to the tendency for the rise of meaning to act as a distraction to the process of cancellation; the latter requires attention to individual letters, whereas the former (ordinary reading) proceeds naturally by the assimilation of words as wholes. As will be noted below, most *S*'s do not 'read' the text when cancelling letters in pied material.

A very good illustration of the effect of form of material upon performance is seen in the comparative results of Brown and Wyatt, both of whom used upon school children 11-12 years old what they refer to as the *er*-test and the *anos*-test. Wyatt's *S*'s cancelled 17.75 *e*'s and *r*'s in 2 min., Brown's cancelled from 362 to 377 in 5 min.; in the *anos*-test Wyatt's *S*'s cancelled 41.20 letters in 3 min., Brown's cancelled 161 in 5 min. The striking disparity is due, so far as one can judge, to the use by Wyatt of 600 letters with only 21 each of the letters to be crossed, and of a page of French words by Brown. All of which shows that, unless a standard procedure is followed, a test cannot surely be recognized by its name.

(6) *Dependence on number of letters to be cancelled.* According to Bourdon, an increase in the number of letters to be cancelled causes a progressive decrease in the extent of material examined, but approximately the same number of letters is cancelled in a given time. This conclusion may be accepted as generally valid, but much depends upon the arrangement of the material (cf. the above comment on Wyatt and Brown). Bourdon is of the opinion that *S*'s that are accurate in cancelling one are also accurate in cancelling several letters. The au-

thor's tests confirm this generalization in the main, as they show a value for this correlation of + 0.38 (Table 42). With regard to speed, Bourdon merely states that some individuals slow in marking one letter prove fast in marking several: my results indicate, however, that the correlation between one- and four-letter efficiency is greater in the case of speed (+ 0.49) than in the case of accuracy (Table 42).

(7) *Relation of speed and accuracy.* Binet, as previously noted, assumes that speed tends to be equalized in marking four letters, accuracy in marking one letter. My tests with 50 boys (Table 41, last column) show that, while there is a tendency in this direction, it is not sufficiently pronounced to warrant the neglect of either speed or accuracy in estimating efficiency. In other words, the variation in speed is proportionately less in marking four letters than in marking one letter, and the variation in accuracy is proportionately less in marking one letter than in marking four letters, but there is nothing approaching equalization of either factor.

TABLE 41

Averages and Variations in Cancellation Tests (Whipple)

LETTERS CANCELLED	INDEX	AVERAGE	M. V.	COEFFICIENT OF VARIABILITY
e	Speed	702	193	27
q, r, s, t	Speed	811	177	22
e	Accuracy	89.6	8.0	9
q, r, s, t	Accuracy	63.1	11.4	18

Cattell and Farrand state that some *S's* are slow and accurate, some slow and inaccurate, some fast and accurate, some fast and inaccurate. Since, however, an *S* that works very fast presumably tends to work less accurately, we may expect to find indications of an inverse correlation between speed and accuracy, and this is the case. Wissler's relatively easy *A*-test gives an inverse correlation ($r = -0.28$). In the author's tests, the cancellation of one letter is harder than that of the *A*-test, and the cancellation of four letters is still harder; in

consequence, we find inverse relations of — 0.37 and — 0.64. respectively (Table 42). In the case of 30 Cornell University students, speed and accuracy were found similarly inversely related, but by a lower coefficient ($r = -0.48$, Table 42).

TABLE 42

Correlations in Cancellation Tests: 50 Grammar-School Boys (*Whipple*)

FIRST MEMBER	SECOND MEMBER	PEARSON COEFFICIENT
Speed, one letter	Accuracy, one letter	—0.37
Speed, four letters	Accuracy, four letters	—0.64
Speed, four letters	Accuracy, four letters	—0.48*
Speed, one letter	Speed, four letters	0.49
Accuracy, one letter	Accuracy, four letters	0.38
Speed, one letter	Class-standing	—0.40
Accuracy, one letter	Class-standing	none
Net efficiency, one letter	Class-standing	—0.32
Speed, four letters	Class-standing	—0.40
Accuracy, four letters	Class-standing	0.39
Net efficiency, four letters	Class-standing	—0.09
Net efficiency, one letter	Word-building (Test 47)	none

*This correlation refers to 30 University students.

(8) *Dependence on movement.* Bourdon's description of the process of cancellation seems to imply that the examination of the line is interrupted during the actual process of cancelling. The same implication is made, perhaps less baldly, in the method used by some investigators to secure a 'perception-time' or 'discrimination-time' by subtracting a 'motor-time' from the total time. The following quotation from MacMillan and Bruner (p. 52) is illustrative:

"It is readily apparent that at least two factors enter into the selection and marking of *a*'s as called for in this test; the one a mental factor concerned in the several acts of discrimination in finding the *a*'s, and the second, physiological, having to do with the motor task of making the marks. To disintegrate these factors and arrive at a measure of the purely mental time actually involved in perception, is the point of chief moment. . . . And this is accomplished approximately by subtracting the *motor time* gotten in the first part of the test from the total time required to perform the second part."

Now, it is perhaps true that the movement of reaction does in some case interfere with the process of perception and recognition, but it is equally true that for others, especially for

adults, there is virtually no interference whatever, and that for other *S*'s the two processes undergo a very considerable degree of 'telescoping.' In any event, the degree of interference may be investigated directly. Thus, Vogt has attempted an experimental analysis of this interference by comparing the amount of ground covered in the usual method and the amount of ground covered when the assigned letters are simply recognized but not marked.[9]

He found that, in his own case, the marking slowed the speed of the performance as a whole by at least 15.8 per cent., in the case of another, less practised *S*, by the astonishing amount, 42.3 per cent. If these conclusions are generally valid, we shall have to admit that the apparently 'mechanical' phase of the cancellation work may easily be the determining factor in the individual differences which the test may reveal, and this will alter radically our conception of the nature of the test.

But, in the author's opinion, Vogt's results are so warped by the intrusion of practise and warming-up (*Anregung*) as to be

TABLE 43

Effects of Different Methods of 'Reaction' in qrst-Test (Whipple)

NUMBER	TIME ELAPSED SINCE PRECEDING TRIAL	METHOD OF REACTING	SPEED IN SEC.	OMITTED OUT OF 199*
1	————	Finger tapping	283	?
2	10 minutes	Electric key	246	3
3	10 minutes	Actual marking	207	3
4	15 hours	Electric key	220	0
5	5 minutes	Actual marking	193	4
6	5 minutes	Mere recognition	197	?
7	2 hours	Electric key	237	1
8	1 hour	Actual marking	184	4

*In the test in use there were only 49 s's, owing to a printer's error.

inconclusive. They certainly do not accord with the results in Table 43, which represent the author's tests upon himself by several variant methods.

In this work the same text and the same letters were employed in eight trials on two days. Test 6 was in accordance with Vogt's series in

[9]There is here no check upon accuracy: it would seem better to let *S* name the letters as fast as recognized, or at least utter some simple sound or tap his finger at each recognition as suggested below.

which *S* merely recognizes the letters, and executes no movement of reaction. In Trial 1, the finger was lifted slightly as each letter was recognized. In Trials 2, 4, and 7, this movement was changed into a simple tap upon an electric key, which was connected to an electric recorder: This device enables one to record the total number of movements of reaction and hence to measure accuracy in terms of omissions, while the movement is so familiar and simple as to be virtually negligible. In the remaining tests, the letters were cancelled by pencil strokes in the regulation manner.

It is evident that when two or three tests are administered in close succession, there tends to be improvement due to practise and warming-up, especially the latter. If the process of marking delayed the performance by any appreciable amount, we should expect Trial 6 to be shorter than Trial 5, and it would be difficult to account for the minimal time shown in Trial 8. Moreover, the author can testify from introspection that there is no conscious delay introduced by the movement of cancelling itself. It is quite possible that some *S*'s may be delayed, however, by the cancelling, and it would be profitable, whenever time permits, to investigate, by appropriate tests, the nature and extent of this retardation in speed, which, if appreciable, will evidently vitiate the test.

(9) *Dependence on practise.* Practise increases efficiency in all cancellation tests, as is illustrated in Tables 40, 43 and 45. Whether this practise-effect concerns only the letter used, or extends to all letters, cannot be stated until a series of equivalent letters has been determined by preliminary tests. Continued practise with the same letters almost doubles speed, and raises accuracy to a maximum. The letters are not held in mind by conscious effort, but recognized quasi-automatically, and the whole process becomes unexpectedly simplified. But the *A*-test, according to Whitley, is little disturbed by practise when a single repetition is made, *e. g.*, time per *A* in the first trial .643 sec., in the second trial .636 sec. On the contrary, repetition of the *a-t* test, even after one or more weeks, "has an effect of over 40 per cent. gain."

The cancellation test has been used by Kent and by Boring as a test of learning, "to test the acquisition of skill in a complicated operation and the effect of skill acquired in one operation upon the acquisition of skill in another very similar operation" (Boring). Both investigators worked at the Government Hos-

pital for the Insane on dementia praecox cases. Boring, whose work was the later and more carefully planned, found that practise under his special conditions (12 days, 5 forms a day) reduced the average time from 47.6 to 33.7 sec., but did not reduce the errors. Contrary to Miss Kent, he found no marked or unambiguous evidence of transfer of practise to new, but equivalent forms.

For a more extended analysis of practise and 'warming-up' gains in the cancellation test see Wells (35).

(10) *Dependence on fatigue.* The experiments of Bourdon, Binet, and Ritter show that cancellation is affected by fatigue, which reduces accuracy, rather than speed—a result in accordance with what we know of the effects of fatigue on other forms of mental activity. It follows that the cancellation of four or more letters is better adapted than the cancellation of one letter for testing fatigue. If but one letter is cancelled, practise and warming-up may easily conceal fatigue, as is shown in the author's continuous tests, extending over two hours and characterized by marked subjective fatigue (Table 40).[10]

In his examination of fatigue, Binet compared the first and second half of a 20 min. test, and found that his *S*'s had made 54 errors in the first, and 95 in the second, 10 min. Ritter was successful in 8 of 10 trials in getting indications of fatigue in school children by his form of the cancellation test: Table 44 gives a sample series of errors.

TABLE 44

Effects of Fatigue on Cancellation (Ritter)

TIME	PREVIOUS SCHOOL EXERCISE	ERRORS
9 a.m.		37
9.55 a.m.	Greek	94
10.10 a.m.	Pause for Vespers	78
12 m.	Livy and Chemistry	84

(11) *Reliability.* Notwithstanding the assertion of Woodworth and Wells that the function tested in number-checking

[10] It may be stated, however, that some three or four tests which followed the fourth group, but which are not included in the table, indicate a distinct increase in time required, *e. g.*, from 85-95 to 106-108 sec. The quality of the work showed no progressive changes.

seems "one whose expression in the test can be distorted by incidental factors that are as yet very imperfectly understood," the fact remains that in nearly all forms of cancellation coefficients of reliability have proved fairly satisfactory: samples are Wyatt for his *er*-test .72, for his *anos*-test .64; Brown for his *er*-test .60 to .97, for his *anos*-test .51 to .56; Winch (see Brown) for his simple motor test (marking all letters) .82 to .91, for his complex motor test (*anos*-test) from .87 to .93; Abelson for his dot-form test .94 to .97; Burt for his *oe*-test .75, and Simpson for the regular *A*-test .60 to .72 and for his geometrical forms test .69 to .91. Simpson believes that four trials might advantageously be used in rating *S*'s with the *A*-test. Whitley concludes that for short tests the *A*-test, *a-t* test and the geometrical forms test are significant of an individual's ability in visual perception.

(12) *Correlation with intelligence.* Earlier investigators, working for the most part by the method of contrasted groups, did not reveal consistent relationship between cancellation and intelligence: later investigators, working by more refined correlation formulas, have usually discovered a small positive correlation.

Binet's results were gained from four different series: in Series 1 (cancellation of five letters), the intelligent and unintelligent showed little difference in speed, though the former increased their speed more in the second half of the test; the unintelligent, however, made four times as many errors. In Series 2, conducted 15 days later, the unintelligent equalled the intelligent in speed, but were still inferior in accuracy. In Series 3, a sudden change was made in the letters assigned; this reduced the speed of all *S*'s more than one-half, and the intelligent, rather oddly, made more errors than the unintelligent. Binet explains this on the ground that they, the intelligent, had established their associations more strongly in previous series.[11] In Series 4 (20 min., 6 letters, two of them new),

[11] These results need reinvestigating. It would be interesting to see whether, in such a test, the intelligent would surpass the unintelligent, provided the latter had, by added practise, been brought to an equal state of proficiency before the change of letters. In accordance with Binet's thesis that intelligence is indicated primarily by readiness to adapt oneself to a new situation, we should then expect the unintelligent to make the greater number of errors.

the speed was about equal for the two groups, but the unintelligent made more errors.

Winteler's gross results (Table 45) indicate the superiority of his intelligent children. but when his *S*'s are ranked individually, two of the four unintelligent are found to be superior to some of the intelligent. Similarly, the unintelligent seemed as capable as the others in adapting themselves to the change from crossing three to crossing five letters, so that Winteler concludes that one cannot, on the whole, discern any inferiority on the part of the unintelligent with respect to the numbers of letters cancelled, to their quickness of adaptation, or to the steadiness with which attention is maintained within the series. It is to be regretted that Winteler did not take any account, direct or indirect, of the number of errors.

TABLE 45

Relation of Average Number of Letters Cancelled to Intelligence (Winteler)

	N S T			L M R S T			
	FIRST DAY	SECOND DAY	BOTH	FIRST DAY	SECOND DAY	BOTH	ALL TESTS
Intell	277	329	606	312	414	726	1326
Unintell	255	303	558	248.5	326.5	575	1133

The author's own tests are summarized in Table 42: it is evident that there is an inverse relation (— 0.40) between speed and class standing; that, when one letter is cancelled, there is no correlation, but when four letters are cancelled, there is a direct correlation (0.39) between accuracy and class standing: that, when accuracy and speed are conjoined in a single index (net efficiency), there is a definite inverse correlation for one letter, and a possible inverse correlation for four letters, between such efficiency and class standing. In other words, the best pupils work more slowly at the cancellation test; if four letters are cancelled, this slower speed has its reward in a relatively high degree of accuracy. The recent work of Lapie (19) lends confirmation to this conclusion: his advanced children

worked with consistent care, his retarded children hastily and somewhat capriciously—which recalls the suggestion made above that good-will and conscientiousness may effect the test with children to an extent sufficient to obscure the rôle of the intellectual factors.

The following are the coefficients of correlation with intelligence reported by recent investigators: Wyatt, *er*-test .37 to .40, *anos*-test .32 to .45; Burt, *oe*-test .39; Brown, *er*-test 0 to .28. *anos*-test .10 to .13; Simpson, *A*-test with 17 highly selected adults .21; Abelson, dot-form test with 88 girls .32, and with 43 boys .28; Descoeudres, with 14 backward children .67—all cases, therefore, of positive correlation, but mostly of rather low degree.

(13) *Correlations with other tests.* Abelson found correlations ranging from .06 to .45 between cancellation and the following tests (averages for both sexes only stated here): interpretation of pictures .20, memory for sentences .23, tapping .26, memory for commissions .17, crossing rings .56, discrimination of length .34, memory of names of objects .12, all these tests combined .41. Wyatt's results, if we may average the correlations for two groups of *S*'s in both his *er* and *anos* tests, are with analogies .32, with the Ebbinghaus test .50, with word-building .43 and with part-wholes .42. Somewhat smaller correlations are reported by the same investigator for a single group of *S*'s with memory for nonsense syllables, interpretation of fables, letter-squares, dissected pictures and Healy's cross-line test. Further work of this sort in which the correlations are less accordant and tend toward zero relationships will be found in Brown. Wissler found no correlation between cancellation and reaction-time to sound, and only a low correlation (0.21) between the *A*-test and a test of quickness in naming colors. The same investigator found that weak eye-sight was conjoined with inaccuracy in the *A*-test, but that the reverse was not true.[12] Thorndike found a correlation between twins

[12] In certain series conducted by the author with a University student to test the relative values of various letters, the net result of two months' work was to indicate the probability that the student had astigmatism!—an inference which was confirmed by the oculist.

of 0.73 in the *A*-test, 0.75 in the *a-t* and *e-r* tests and in the misspelled word test.

The author's grammar-school boys showed no correlation between net efficiency in cancelling one letter and the word-building test.

Aikins, Thorndike and Hubbell found no correlation between quickness in the misspelled word test and quickness in the *e-r* test, and a correlation of 0.16 (8th-grade pupils) to 0.25 (5th-grade pupils) between accuracy in these two tests, as measured by number of words or letters marked per line. The correlation of efficiency in the *e-r* test and in addition and association tests was also found to be slight or none, but efficiency in misspelled words and in addition and association tests was correlated by 0.50.

Heymans and Brugman (15a) found but a low correlation (0.16) between the cancellation of numbers containing three specified digits and two other tests of 'concentration'—the McDougall dotting test and listening to faint noises made by a fall phonometer. They conclude that the lack of correspondence was due to certain difficulties in the use of the two tests that were compared with cancellation.

The cancellation test has also been used by Wallin (see his Table V, pp. 26-7) and by Kohnky, in conjunction with other tests, for measuring mental improvement due to dental inspection in public-school classes, and by Hollingworth in his investigation of the influence of caffein. In the caffein experiment the general effect is stimulation for large doses and retardation for small, and the stimulation is more evident in *S*'s that work by recognition and without perceptual discrimination: these conclusions, however, are given with considerable reservation, because the test was found not well adapted for this work.

(14) *Defectives.* Miss Norsworthy used the *A*-test and the *a-t* test upon 150 feeble-minded children. Compared with normal children at three points in the surface of distribution, median, above — 1 *P. E.* and above — 2 *P. E.*, corresponding to 50, 75 and 91 per cent., respectively, the frequency for feeble-minded children was found to be 9, 18, and 34 for the *A*-test and 1, 14, and 28 for the *a-t* test, *i. e.*, 18 per cent. of the feeble-

minded reach a score attained by 75 per cent. of normal children, etc. The results with Vineland children are reported in Table 39. It will be seen that there is general improvement with mental age, but Doll remarks that the mean variation and the extremes of performance are both large; moreover, "at no age do 80 per cent. or even 75 per cent. surpass the average of 80 per cent. of the cases at the preceding age," so that the test has obvious defects as a diagnostic test. Another defect is found in the lack of correspondence between repeated examinations of the same feeble-minded child.

Tests by MacMillan and Bruner with pupils attending the public day-school for the deaf at Chicago show (1) that, taking all cases (about 140) from 7 to 16 years, an average of 82.5 per cent. of these children are slower than hearing children in perception-time, (2) that, roughly speaking, motor-time and perception-time are about the same at all ages. These authors conclude that "deaf children, on the average, are from two to three years less mature than hearing children of the same ages in the kind of ability that is called for in this test."

Strong says that it "seems very probable that both depressive and manic attacks interfere with performance in the cancellation test" (by increasing the time).

Fernald tested out and retained the cancellation test in his effort to arrange a group of diagnostic tests to use with defectives and delinquents.

(15) *Qualitative analysis.* Most *S*'s do not pronounce the letters of the text, as is shown both by introspection and by the fact that a greater number of letters can be examined than could be read over silently in the same period, *e. g.*, 1876 and 1086, respectively (Bourdon). The letters to be cancelled, however, are often mentally pronounced by *S*, especially if four in number, in order to keep them in mind. Attention is then arrested by the sight of the assigned letters, which are recognized either visually or by inner pronunciation. (Cf. the reference previously made to Meumann's analysis.)

The most common error is that of omission. When four letters are marked, *S* often temporarily forgets one or more letters. Less often, *S* makes no marks at all for limited parts of a

line, or even for whole lines. The latter defect is, in the author's experience, characteristic of either very young, or very careless *S*'s.

Adults may hit upon the device of traversing every other line from right to left; this seems to economize time and to insure at least as great accuracy.

REFERENCES

(1) A. R. Abelson, The measurement of mental ability of 'backward' children. *BrJPs*, 4: 1911, 268-314. (Same material in simpler form, Tests for mental deficiency in childhood. *The Child*, 3: 1912, 1-17.)

(2) H. Aikins, E. Thorndike, and E. Hubbell, Correlations among perceptive and associative processes. *PsR*, 9: 1902, 374-382.

(3) A. Binet, Attention et adaptation. *AnPs*, 6: 1900, 248-404.

(4) A. Binet, L'étude expérimentale de l'intelligence, Paris, 1903. Pp. 309.

(5) E. G. Boring, Learning in dementia praecox. *PsMon*, 15: 1913, No. 63. Pp. 101, especially 30-50.

(6) B. Bourdon, Observations comparatives sur la reconnaissance, la discrimination et l'association. *RPhF*, 40: 1895, 153-185.

(7) W. Brown, The essentials of mental measurement. Cambridge, 1911. (See also *BrJPs*, 3: 1910.)

(8) W. Brown, The effects of 'observational errors' and other factors upon correlation coefficients in psychology. *BrJPs*, 6: 1913, 223-238. (Including a section by W. H. Winch, on Description of simple motor and complex motor tests, pp. 228-230.)

(9) C. Burt, Experimental tests of higher mental processes and their relation to general intelligence. *JEPd*, 1: 1911, 93-112.

(10) J. Cattell and L. Farrand, Physical and mental measurements of the students of Columbia University. *PsR*, 3: 1896, 618-648.

(11) W. G. Chambers, Individual differences in grammar-grade children. *JEdPs*, 1: 1910, 61-75.

(12) Alice Descoeudres, Exploration de quelques tests d'intelligence chez des enfants anormaux et arriérés. *ArPs (f)*, 11: 1911, 351-375.

(13) E. A. Doll, The A-test with the feeble-minded. *TrSc*, 10: 1913, 49-57.

(14) G. G. Fernald, The defective-delinquent class: differentiating tests. *AmJIns*, 68: 1912, 523-594.

(15) S. I. Franz, Handbook of mental examination methods. *JNeMeDisMon*, No. 10. 1912. Pp. 165, especially 129ff.

(15a) G. Heymans und H. Brugmans, Intelligenzprüfungen mit Studierenden. *ZAngPs*, 7: 1913, 317-331.

(16) H. L. Hollingworth, The influence of caffein on mental and motor efficiency. *ArPs(e)*, No. 22, 1912. Pp. 166.

(17) C. Judd, Laboratory manual of psychology. N. Y., 1907. Pp. 124.

(18) Emma Kohnky, Preliminary study of the effect of dental treatment upon the physical and mental efficiency of school children. *JEdPs*, 4: 1913, 571-8.

(19) P. Lapie, Avancés et retardés. *AnPs*, 18: 1912, 233-270.

(20) D. MacMillan and F. G. Bruner, Special report Dept. of Child-Study and Pedagogic Investigation on children attending the public day-schools for the deaf in Chicago. 1908. Pp. 88.

(21) E. Meumann, Vorlesungen zur Einführung in die experimentelle Pädagogik. 2d ed., Vol. 2, Leipzig, 1913, especially 393 f.

(22) H. Münsterberg, Psychology and industrial efficiency. Boston, 1913. Pp. 321. Especially 77 ff.
(23) Naomi Norsworthy, Suggestions concerning the psychology of mentally deficient children. *JPsAsth*, 12: 1907, 3-17. (See also her Psychology of mentally deficient children, New York, 1906. Pp. 111.)
(24) A. Oehrn, Experimentelle Studien zur Individualpsychologie. *PsArb*, 1: 1896, 92-151.
(25) C. Ritter, Ermüdungsmessungen. *ZPs*, 24: 1900, 401-444.
(26) E. C. Sanford, The relative legibility of small letters. *AmJPs*, 1: 1888, 402-435.
(27) Stella Sharp, Individual psychology: a study in psychological method. *AmJPs*, 10: 1899, 329-391.
(28) B. R. Simpson, Correlations of mental abilities. *Columbia ConEd*, No. 53, 1912. Pp. 122.
(29) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, 430-443, 493-506, especially 370-374 and 500 ff.
(30) E. K. Strong, Jr. A comparison between experimental data and clinical results in manic-depressive insanity. *AmJPs*, 24: 1913, 66-98.
(31) E. L. Thorndike, An introduction to the theory of mental and social measurements, N. Y., 1904. Pp. 210.
(32) E. L. Thorndike, Measurements of twins. *ColumbiaConPhPs*, 13: No. 3. Pp. 64.
(33) R. Vogt, Ueber Ablenkbarkeit und Gewohnungsfähigkeit. *PsArb*, 3: 1901, 62-201.
(34) J. E. W. Wallin, Experimental oral euthenics. *Dental Cosmos*, April and May, 1912. Reprint, 32 pp.
(35) F. L. Wells, (*a*) The relation of practise to individual differences. *AmJPs*, 23: 1912, 75-88. (*b*) Practise and the work-curve. *AmJPs*, 24: 1913, 35-51.
(36) Mary T. Whitley, An empirical study of certain tests for individual differences. *ArPs(e)*, No. 19, 1911. Pp. 146, especially 61-75.
(37) J. Winteler, Experimentelle Beiträge zu einer Begabungslehre. *EPd*, 2: 1906, 1-48, 147-247.
(38) C. Wissler, The correlation of mental and physical tests. *PsMon*, 3: No. 6, 1901. Pp. 62.
(39) R. S. Woodworth and F. L. Wells, Association tests. *PsMon*, 13: 1911, No. 57. Pp. 85.
(40) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

TEST 27

Counting dots.—This test was devised by Binet (1) and employed by him and later by Winteler (2) in their comparative studies of intelligent and unintelligent children in order to measure the degree of attention. The problem set before *S* is that of counting a number of dots which are arranged in an irregular group or in lines of varying length and spacing.[1]

[1]The test evidently has similarity with the dot-counting under 3 sec. exposure (Test 25), but the removal of the time restrictions and the increase in the number of dots make the conditions quite different: here it is the degree rather than the range of attention that is primarily to be measured.

When this work is attempted without the aid of pointer or pencil to keep the place, it is distinctly difficult and necessitates active concentration, but by selecting different arrangements of dots, this difficulty can be graded to suit S's of different degrees of development, and comparative scales of ability or norms for given arrangements of dots may presumably be established for each age.

MATERIAL.—Stop-watch. Two sets (in duplicate) of 27 printed test-cards.

These cards are numbered in the upper left-hand corner in accordance with the following plan: Cards A 1 to A 10 contain rows of dots with uniform spacing (for each card): Cards Ba 1 to Bd 4 contain lines with groups of 2, 3, 4 or 5 dots each in which the spacing within the groups and between the groups varies as indicated in Table 46. Cards C 1 to C 5 contain 5 arrangements of dots in irregular clusters. These three kinds of material reproduce those found to be of value by Binet and Winteler. The dots, like Winteler's, are 1.5 mm. in diam.[1] The 'A' cards were used by both experimenters, the 'B' cards by Winteler only, the 'C' cards by Binet only. The term 'points' in the Table refers to the printer's point or typographic unit: one point is 1/72 inch.

TABLE 46

Specifications for Test-Cards Used in Dot-Counting.

CARD NUMBER	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Points in Spacing	20	19	12	10	8	5	4	3	3	3
Number of Dots	13	15	14	17	21	27	29	45	52	60

CARD NUMBER	Ba1	Ba2	Ba3	Ba4	Bb1	Bb2	Bb3	Bb4
Points within Groups	6	6	6	6	4	4	4	4
Points between Groups	18	18	18	18	12	12	12	12
Number of Dots	45	47	55	52	50	54	70	64

CARD NUMBER	Bc1	Bc2	Bc3	Bc4	Bd1	Bd2	Bd3	Bd4
Points within Groups	3	3	3	3	3	3	3	3
Points between Groups	9	9	9	9	15	15	15	15
Number of Dots	70	73	81	91	65	75	70	72

Card Number and Number of Dots: C1, 50; C2, 48; C3, 49; C4, 54; C5, 57.

[1]Binet found that S's with poor eye-sight had difficulty with the test when the dots were 1 mm. in diam.

METHOD.—Make the tests in the order indicated by the card-numbers. Instruct *S*: "Find the total number of dots on this card: count aloud, in any way you wish (that is, by ones, twos, threes, etc.): work as rapidly as you can, but try particularly to get the number right." The emphasis is thus placed on accuracy rather than upon speed. *S* must not use his finger, pencil or other object to keep his place. *E* should record *S*'s time, but without the latter's knowledge. He should also keep before him a duplicate of the card upon which *S* is working and should note thereon *S*'s method of grouping, in order to discover whether he counts always by ones, or always by twos, etc., and whether he accommodates himself to the objective grouping of the 'B' cards. (*S*'s attention should not, of course, be called by *E* to the grouping in these cards.) *E* should also record the magnitude of the error and its nature—whether overestimation or underestimation—but should not communicate this information to *S*.

TREATMENT OF RESULTS.— Binet and Winteler both ranked *S*'s merely in terms of accuracy and put no time-limit upon their performance. It would seem possible, after some experience, to discover the relation between speed and accuracy, and possibly to make use of a corrective formula as in the case of other tests where these two factors appear, *e.g.*, the Cancellation Test (No. 26). The errors are to be counted simply by subtracting the given from the true number, *e.g.*, 62 for 65 represents 3 errors.

RESULTS.—(1) Many *S*'s have a *constant error*, but this may be either an error of overestimation or an error of underestimation.[3]

(2) Winteler found that some *S*'s always counted by the same number, *i.e.*, used the same *increment in adding*, as "one, two, three, four," etc., or "two, four, six, eight," etc. When

---

[3]No attempt has been made to analyze the conditions under which these constant errors appear; they might conceivably be due in part to illusions of filled and empty space, in part to individual differences in method of keeping the place in the line, in part to temperamental attitudes (over-cautiousness, careless haste, etc.). This test, like many others, has not been subjected to careful introspective analysis by trained adults. But critical qualitative study of this sort is as desirable for the intelligent employment of any test as is the mere accumulation of quantitative results.

the change was made to the 'B' material, he found that all three of his bright children and one of the dull adapted their counting to the objective grouping, whereas the rest of the dull children continued for the most part to employ the form of counting (almost invariably by ones or by twos) that they had adopted in the 'A' cards.

(3) In examining the *relation between dot-counting and intelligence,* Binet concludes that, although the test exacts a high degree of attention, the outcome depends more upon *S*'s care than upon his intelligence. His results were confused by the presence of one bright child with poor eye-sight: when this case is eliminated, the intelligent children are found to make fewer errors than the unintelligent: in one series the relation is 13 to 19, in another 16 to 24; the difference, as in many other tests, tends to lessen with practise.

Winteler likewise found the bright children, as a group, more accurate, and also more rapid, than the dull children, but there were individual exceptions. The average for 10 series on each of two days were: errors, bright 8.33, dull 17.25: time in sec. per 10 dots, bright 4.97, dull 6.30—a suggestion that the tendency to inverse relation of speed and accuracy may not be so clearly evident as in many other tests. The difference in capacity of the two groups becomes striking, however, when the 'B' type of material is employed, because the unintelligent, as already noted, fail to adapt themselves to the objective grouping and make a large number of errors (79.75 to 9.00 of the intelligent).

(4) The outcome of the test is not affected by general ability in arithmetic; some of Winteler's *S*'s who did poorly had good grades in arithmetic.

REFERENCES

(1) A. Binet, Attention et adaptation. *AnPs,* 6: 1899 (1900), 248-404.

(2) J. Winteler, Experimentelle Beiträge zu einer Begabungslehre. *EPd,* 2: 1906, 1-48, 147-247.

TEST 28

Reading simple and complicated prose.—Reading, like counting, is a form of perceptual activity that has been proposed as a means for measuring attention. In reading, as in counting (Test 27), the process has been complicated in some man-

ner in order that the increased difficulty may exact a higher degree of attention and so furnish a better opportunity for the study of individual differences. Miss Sharp, for instance, followed this plan when she sought to test degree of attention by requiring *S*'s to read two texts (a page of concrete description and a page of abstract exposition), which were printed without capitals, punctuation, or spacing.

In the present test, this plan has been extended, first by printing the complicated text backward, as well as without spacing, and second, by adding a test-sheet of similar subject-matter and identical length, but of regular form, in order to supply a check-test of maximal speed of reading under normal conditions.

This simpler test of reading under normal conditions may also be used with immature *S*'s as a direct index of general mental status. It was in this form, for example, that Binet and Simon used reading in their 1908 scale as a differentiating test between imbecility and moronity. Wallin's work with epileptics shows that the test has decided value in this direction, so that its elimination from the Binet scale in 1911 seems at least unfortunate.

MATERIALS.—Two printed texts: (*a*) a page of prose in regular form, (*b*) a page of equivalent, but 'complicated' prose. Stop-watch.

METHOD.—The regular text (*a*) is used first. *E* gives the following directions. "When I say 'now,' I want you to read this aloud, just as fast as you can without making mistakes." With older children and adults the tendency to slurring and skipping must be checked by the further direction to "read clearly enough so that another person who did not know the passage could understand you." *E* records the time; also, if desired, the number of errors. He does not, however, correct *S*'s errors in this part of the test.[1]

[1]With most adults the errors are few in number and trivial in nature. To attempt to correct them would render it impossible to measure the speed of reading. If the errors are numerous, *E* may, however, ask *S* to reread the text, this time without making a single error, but still as fast as possible. The advantage that *S* gains by knowing the subject-matter is of little moment compared to the false advantage that he has gained by hurrying his reading so fast as to commit many errors.

In using the reversed and unspaced text (*b*), *E*'s directions are: "When I say 'now,' read this aloud as fast as you can without making mistakes. You will find this page more difficult than the one you have just read, because you will have to begin here in the lower right-hand corner and read it backward, and because this page is printed without any punctuation or capitals and without spaces between the words. I shall not give you any help, but if you make a mistake, I shall stop you, and ask you to correct it."

E follows *S*'s reading upon a duplicate text. He records the time for the entire reading, and by glancing at the watch at every pause in the reading, he notes upon the duplicate text at the points concerned, the time in sec. consumed by *S* at these pauses. These notations should be made for every pause of 5 sec. or over. In case *S* pauses for 30 sec. at any point, *E* then supplies for him the word or phrase which he needs to continue his work.

To secure accuracy, *E* must correct every error in *S*'s reading, even slight errors, such as singular for plural forms, etc., and he must especially avoid the temptation to assist *S* whenever he halts, save for the 30 sec. halts, as just stated. He must notify *S* of each error, as it occurs, by simply interjecting 'no,' and must indicate its place by repeating the two or three words just preceding it. For example, if *S* reads: "they were all alike in tone," *E* interrupts with: "No!—all alike in?—" *S* corrects himself: "all alike in one respect," etc.

VARIATIONS IN METHOD.—In order to measure, and to be able to allow for, individual differences in maximal rapidity of articulation, *E* may require *S* to reread the normal text four or five times, or until the subject-matter is thoroughly familiar and further repetition fails to reduce the time. A brief rest-pause should follow each trial to avoid cumulative fatigue. Another method is to test *S*'s time for counting aloud to 50 at maximal speed, though here it is often more difficult to check or control *S*'s tendency to gain time by slurring or otherwise mutilating the words he pronounces.

In view of the correlation mentioned below between speed of reading forward and speed of reading backward, it is of in-

terest to ask S after his first reading what it was that set the limit to his speed. The usual answer—inability to articulate faster—appears not wholly adequate in view of the fact that many S's can read the passage slightly faster on a second trial. To insure that this gain is not due to simple 'warming-up' of the vocal apparatus, E might give S a preliminary exercise with some other text.

RESULTS.—The results obtained by Burt with two groups of school children at Liverpool with the simple task of reading forward are shown in Table 47, those obtained by the author with the texts prescribed above are shown in Table 48. From these data the conclusions that follow are drawn.

TABLE 47

Results of Reading Tests at Liverpool (Burt and Moore).

Kind of Reading	Ages	Mean or Median for Boys	Mean or Median for Girls	Coefficient of Reliability	Correlation with Int'lig'ce	Excess Perc'tage for Boys
Aloud	12.5	224.2 min.	191.5 min.	--	.26	---
Silently	12.5	181.5 min.	138.8 min.	--	.21	---
Mass test	13	231 words	315 words	.46	.00	—32.7
Silently	13	166.0 sec.	134.6 sec.	.69	.22	—32.1
Aloud	13	211.8 sec.	186.4 sec.	.81	.29	—36.6

TABLE 48

Results of Reading Tests, in Sec. (Whipple).

Group	Number	Reading Forward Av.	Reading Forward M. V.	Reading Forward Min.	Reading Forward Max.	Reading Backward Aver.	Reading Backward M. V.	Reading Backward Min.	Reading Backward Max.
Dull Children	5	116.4	19.	101	169	1061.	291.	814	1500*
Bright Children	5	100.0	11.6	85	125	544.4	167.2	490	910
University Students	26	73.3	8.5	47	100	320.	100.	125	755

*This time is estimated from the amount of the text covered in 10 minutes.

(1) *Dependence on age.* Speed of reading forward increases with age. The average for five bright children aged 10.33 to 12.75 years is less than that of university students, though the fastest child excels the slowest adult. The same

relations obtain for speed in reading backward, though here the increased difficulty entails a still greater difference between children and adults.

Wallin tested epileptics at Skillman, N. J., with the aid of the passage of prose containing 53 words that formed a portion of the Binet-Simon 1908 scale. His averages (in sec.), in comparison with the standards laid down by Binet and Simon for normal children, all expressed in mental ages, may be summarized as follows:

MENTAL AGE	VII	VIII	IX	X	XI	XII
Average according to B.-S.	----	45	40	30	25	---
Average for epileptics	129.2	86.5	61.9	44.6	26.5	23.2

There is, then, a gradual increase in speed of reading with age, but epileptics of a given mental age average slower than normal children of the same mental age.

(2) *Dependence on sex.* All observers agree that as a class females read faster than males. The early study of Romanes (1887) showed a distinct superiority of women over men, while the sex differences calculated by Burt and Moore all range between 32 and 36 per cent. in favor of the girls, *i.e.,* about two-thirds of the girls exceed the median performance of boys.

(3) *Individual differences* are large in either form of test, and are specially accentuated when the reversed text is used.

The coefficient of variability, for example, in the group of adults is approximately 11 per cent. for forward, as contrasted with 33 per cent. for backward reading.

The actual range of performance is also surprisingly large. Thus, in the adult group, the quickest backward reading is 1/6 as long as the slowest backward reading. Miss Sharp's seven *S*'s showed even greater individual differences—ranging from 143 to 900 sec. for concrete, and from 125 to 405 sec. for abstract texts. When it is remembered that her *S*'s were all college students in advanced classes, the variability in performance seems unexpectedly large, and it is hard to understand Miss Sharp's declaration: "We had expected to discover individual differences of much more definite character and much greater amount."

(4) *Correlation with mental ability.* A comparison of the performance of the dull and the bright children of approxi-

mately the same school grades shows the clear superiority of the bright pupils, despite the fact that they are some two years younger than the dull pupils. Here, again, the difference is accentuated in the complicated text. The more extensive tests of Burt also reveal a direct, though slight correlation (about .25) between speed of reading forward and estimated intelligence.

(5) *Relation of speed reading forward to speed reading backward.* The author's tests show a correlation of .79, *P.E.* about .09, between the rate of reading the normal text and the rate of reading the complicated text.

This relation appears, to the author at least, unexpected, and hence of special interest. Adults who try the test are almost unanimous in their declaration that their speed in reading forward is not conditioned by the task of assimilating the substance of the text, but solely by the physiological limit to intelligible articulation: their speed in reading the reversed text, however, is quite obviously not conditioned by speed of articulation, but by a sort of 'linguistic readiness,' or ease of apperceiving the constituent words or phrases of the text. If these statements are correct, we are evidently driven to the conclusion that persons who read difficult and complicated subject-matter rapidly also tend to speak more rapidly—a conclusion that subsidiary tests appear to confirm. It is, furthermore, not unlikely that fast readers are also fast thinkers as well as fast speakers, though this generalization has at present no experimental verification.

NOTES.—We have not had sufficient experience as yet with the complicated, or reversed prose test to understand fully the nature of the processes upon which it depends. The considerations just developed make it evident that these processes embrace something besides attention, if, indeed, attention plays any large share in the conditioning factors. For fast readers, in the reversed text, the proper combinations 'rise up' like the hidden faces in the puzzle picture when once they have been seen.

Since a facile apperception of printed symbols would appear, on theoretical grounds, to be a natural concomitant of good intelligence, it is possible that this test may prove to have considerable diagnostic value. To determine this point, however, it needs extended trial with control both by introspective analysis and by the statistical examination of all possible functional correlations.

REFERENCES

(1) C. Burt, Experimental tests of higher mental processes and their relation to general intelligence. *JEPd,* 1: 1911, 93-112.

(2) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd,* 1: 1912, 273-284, 355-388.

(3) G. Romanes, Mental differences between men and women. *Nineteenth Century,* May, 1887.

(4) Stella E. Sharp, Individual psychology: a study in psychological method. *AmJPs,* 10: 1899, 329-391.

(5) J. E. W. Wallin, Experimental studies of mental defectives. *EdPsMon,* No. 7: 1912. Pp. 155.

TEST 29

Simultaneous adding.—In Tests 24 and 25, the attempt was made to measure the field of consciousness or range of attention during a relatively brief period. Tests have been proposed in which attention is solicited by several claimants, not for a brief period, but continuously. These tests may be grouped as tests of simultaneous activity. Their primary purpose is to ascertain how successfully a number of activities can be carried on at the same time. Ordinarily, disparate activities are selected (Test 30). Less often, as in the present case, a single type of activity is complicated or 'spread' in such a manner as to demand simultaneous attention to more than one phase of activity.

In simultaneous adding, as devised and conducted by Binet in his comparative study of six bright and six dull children, the task is to carry on a series of additions in three columns of figures at once.

Materials.—Prepared forms, ruled in series of three vertical columns, with the numbers 6, 28, 43, printed at the head of the first three columns. Stop-watch. Pencil. A piece of cardboard about 20 cm. square.

Method.—Explain to *S* the arrangement of the form and its three columns. Let him glance for a few seconds at a sample form which has been previously filled out. Make clear to him that he is to continue for six minutes, as fast as he can, to add *one* to each number and to write the sum directly below the line just written. Thus, he first writes: 7, 29, 44, then 8, 30, 45, etc. Inform him that the moment he has written a line (three

sums), this line will be covered with the cardboard, so that he must hold the three sums in his mind from one line to the next. Make clear also that when the foot of one vertical column is reached, he must proceed without delay to fill out the next column. Finally, warn *S* that in case he at any time forgets entirely the sum that he has just written, he must guess at it and continue the work as well as he can.

Record the number of additions made in 6 min., and note the number and nature of the errors. A convenient method is to run the eye vertically down each column and mark a ring around each number that is not an addition of one to the number just above it.

VARIATIONS OF METHOD.—For mature *S*'s, the addition of 1 to each sum may not be difficult enough; *E* may then complicate the task by requiring *S* to add, say, 3 to the first, 1 to the second, and 2 to the third column. If desired, repeat the test with another variation in the constants to be added, say, 2 to the first, 3 to the second, and 1 to the third column.

RESULTS.—(1) Binet found that this test excited a great deal of interest, and exacted a greater effort of attention than any he tried.[1]

(2) There are marked differences in the *amount of work* (number of additions) done by different *S*'s, *e.g.*, in Binet's tests, from 40 to 96 numbers, *i.e.*, from 13 to 32 lines of 3 numbers each.

(3) Examinations of the *errors* shows that *S*'s commonly center their attention upon either one or two columns: here they make few or no errors, but the neglected columns contain many errors. In other words, errors in all three columns at once are rare. Binet's pupils wrote 100 of 245 lines erroneously; of these 100 errors, 59 were found in one column only, 34 in two columns, and 7 in all three columns. Errors are more frequent in the third column than the second, and in the second than the first; by count, 76, 60 and 31, respectively.

There is often displayed a tendency toward a form of error that (unintentionally) simplifies the relation between the col-

[1] For detailed illustrative work secured from his *S*'s, consult his article, p. 384.

umns, *e.g.*, the difference between the second and third columns suddenly drops to 10 instead of 15. Once such an error is made, the chance is greatly increased that *S* guides his work by the difference from column to column, and the test is practically invalidated. Furthermore, the earlier such an error occurs, the more *S*'s performance is facilitated and his score increased. For this reason the variant method suggested above, whereby the differences between the columns are constantly shifting, is much to be preferred for testing mature *S*'s.

(4) Contrary to the results of many laboratory tests, says Binet, the number of errors committed tends to increase with continued work. This, we may surmise, is due in part to the fact that the later portion of the work necessarily deals with higher numbers, and in part to the confusion and loss of confidence that is felt after a number of errors have been made.

(5) *Degree of intelligence* appears to have little influence upon this test. Binet states that speed is not at all a matter of intelligence, merely an individual variation. The unintelligent make more errors than the intelligent, but the difference (17 *vs.* 13.4) is not as great as one might expect.

NOTES.—As a matter of fact, it is improbable that simultaneous adding really exacts a 'spread' of attention to all three columns. What *S* does is to write a given line on the basis of his memory of the preceding line: his additions are made successively; he is not really adding three columns at once. The test, therefore, really measures what is termed "immediate memory" rather than the spread of attention. The results may also be conditioned by the readiness with which *S* handles numbers and performs additions generally.

REFERENCE

A. Binet, Attention et adaptation. *AnPs*, 6: 1899 (1900), 248-404.

TEST 30

Simultaneous disparate activities.—In tachistoscopy and visual apprehension (Tests 24 and 25), we measure the range of attention for simultaneous *impressions*: in simultaneous adding (Test 29), we test the capacity of attention for concurrent

activities of a *homogeneous* type: in the present test we employ *disparate activities,* and study what Meumann terms "heterosensory" distribution of attention (8, i., p. 503). Theoretically, the measure of the capacity of an individual to direct his attention to the execution of several activities at once should be of importance, since this capacity seems to imply the possession of such traits as keen concentration, mental alertness, quick-wittedness, and general intelligence. Gifted men, like Napoleon and Caesar, are said to have possessed this capacity in high degree. The latter, for instance, could dictate four letters while writing a fifth.

There seems to be a possibility that such a distribution of attention may take place under some circumstances, at least a distribution to two lines of activity, but strict experimental examination of the phenomenon is not easy, because, in practise, one of two things usually occurs; if both activities are difficult, attention alternates between them, and the activity not attended to at any moment is temporarily reduced, if not altogether suspended; and if one activity is relatively easy, it becomes, after short practise, reduced to automatism, so that attention can be given freely to the other. It is, then, difficult, if not impossible,[1] so to arrange experimental conditions as to secure *continuous division* of attention to disparate activities. On the other hand, it may be said that the very capacity to alternate attention quickly and successfully from one activity to another, or to reduce one activity quickly to automatism, is itself an indication of important capacities—particularly of well-trained, highly concentrated attention, quick adaptability, and general mental alertness.

The test selected, simultaneous reading and writing, is but one of a large number of possible combinations, others of which are suggested below. This test has been proposed by Paulhan (9), by Binet and Henri (2, pp. 446-7), and tried in several forms by Miss Sharp (10) as a test of "range of attention."

Materials.—A selected poem—preferably one which is divided into a number of stanzas of equal length—which will be

[1] Cf. the remark of Titchener (12, p. 376): "Simultaneity of two psychologically disparate 'attentions' is, in my experience, altogether impossible."

of interest to *S*, but which is not well-known to him. Stop-watch. Pencil and paper.

METHOD.—(1) Let *S* read aloud at his *normal rate* a given section (about 8 lines) of the poem.

(2) Let *S* read another section[2] of the same length, and while reading, write the letter *a* as many times as possible. Continue the test by the use of other sections of the poem combined with the writing (3) of *a b*, (4) of *a b c*, and (5) of the entire alphabet.

This last test is the most satisfactory and should be the one employed if time permits but a single trial. It is important that *S* should *try* to maintain his reading at the normal rate.

VARIATIONS OF METHOD.—(1) Let *S* read both passages at his maximal instead of at his normal rate.

(2) Repeat the test several times with fresh texts to determine the effect of practise upon its performance.

(3) Compare the effect of striving especially for a large number of letters written with the effect of striving especially for a normal or for a maximal rate of reading.

TREATMENT OF RESULTS.—In practise, it will rarely be found that *S* maintains his normal rate of reading, particularly when writing the whole alphabet. To avoid the difficulty of working with two quantities, rate of reading and number of letters written, it is desirable to reduce these to a single "index of simultaneity." This is done, as is illustrated for the whole alphabet test in Table 49, by subtracting the normal reading-time from the reading-time during simultaneous activity, and dividing the number of letters written by this difference.[3] Thus, ob-

[2]Sharp used prose, and had her *S*'s read the same section five times. This has the disadvantage of tending to automatic reading. The advantage of securing identical length is practically assured here by the use of stanzas of poetry.

[3]Miss Sharp divided the difference by the number of letters. The reverse procedure has the advantage of indicating the degree of simultaneity directly, as a large quotient means good ability. Whichever method is used for figuring the index of simultaneity, it is difficult to avoid a certain arbitrariness by any fixed formula. Possibly a more satisfactory formula could be worked out empirically on the basis of a considerable mass of data. The methods employed by Simpson in another connection (Correlations of mental abilities. *Columbia Univ. Contr. to Educ.*, No. 53, may be consulted for suggestions to this end.

server *B* read normally in 28 sec., with alphabet-writing in 113 sec.—a difference of 85 sec. He wrote 91 letters, and has an index of 1.07.

RESULTS.—The results for Miss Sharp's seven *S*'s are summarized in Table 49. It will be noted, (1) that the reading-time is usually lengthened by the complication of writing, (2) that more letters can be written with three than with two letters and more with two letters than with one letter, but not fully three times and two times as many, (3) that the writing of the whole alphabet is much different; either the reading is very much slowed, or fewer letters are written than when only *a*

TABLE 49

Simultaneous Reading and Writing (Sharp).

S's	TIMES OF THE FIVE READINGS IN SECONDS					NUMBER OF LETTERS WRITTEN				Reduction in 5th Reading	Index of Simultaneity 5th Test
	1st	2d	3d	4th	5th	*a*	*ab*	*abc*	Alph'b't		
B.	28	38	42	50	113	47	62	78	91	85	1.07
G.	22	22	22	21	28	29	34	39	40	6	6.66
V. M.	29	30	30	30	50	40	56	57	46	21	2.19
W. M.	26	27	27	27	29	27	28	36	13	3	4.33
E. R.	27	27	29	27	31	31	40	48	20	4	5.00
L. R.	22	25	26	25	37	41	44	51	26	15	1.76
T.	27	29	30	31	29	36	40	45	25	2	12.50
Average	26	28	29	30	45	36	43	51	37	19	4.78

is employed, (4) that the *S*'s differ markedly in their capacity to carry on two processes simultaneously. The rank of the *S*'s in this test did not, however, correlate with their rank in any other of Miss Sharp's tests.

(5) Paulhan noted that the simultaneous performance of two relatively easy activities did not take as long as the performance of the two in succession. He says: "I write the first four verses of *Athalie*, whilst reciting eleven of Musset. The whole performance occupies 40 sec. But reciting alone takes 22 and writing alone 31, or 53 altogether, so that there is a

difference in favor of the simultaneous operations.[*]" And again: "I multiply 421,312,212 by 2; the operation takes 6 sec.; the recitation of 4 verses also takes 6 sec. But the two operations done at once only take 6 sec., so that there is no loss of time from combining them."

NOTES.—Several *other tests* of a similar nature may be briefly described; still others may be contrived by *E* to suit conditions.

(1) As suggested by Meumann (p. 504), the *Cancellation Test* (No. 26) may be combined with other forms of activity, *e.g.*, let *S* cancel one or more letters and at the same time repeat short sentences read to him by *E*, or listen to the reading of a page of narration (Cf. Test 39) and repeat as much as possible of it after the cancellation is finished, or discriminate two points on the skin (Test 23), etc.

Vogt (13) combined the cancellation of three letters in a nonsense test with reaction to *metronome-beats* in the following manner: the metronome was set at 38, and the bell attachment set for every other stroke, so that there were 19 bell-strokes per minute; in some series *S* was required to make a slight movement of the finger at every bell-stroke, in other series also to lift two fingers at every fourth bell-stroke. Vogt found that this 'metronome-counting' retarded the total process of cancellation from 11.6 per cent. to 35.2 per cent., but that it did not affect appreciably the simple apprehension of the letters without actual cancellation (see Test 26, Result 8); in other words, he concludes that the movements of reaction to the metronome interfered with the movements of reaction in cancelling, but did not interfere with the apprehension of the letters in cancelling. This result is difficult to interpret if we do not admit Vogt's contention that the marking is in itself an appreciable factor in the cancellation test.

(2) McDougall (7) has proposed a form of 'dot tapping' to test the capacity for continuous exertion of attention in the following manner: Place upon a kymograph drum a sheet of white paper on which have been printed eight rows of 120

[*]The 'telescoping' may amount to more than this in the case of some individuals. See Titchener, 12, p. 375.

red dots; each dot is 1.5 mm. in diameter, and 5 mm. distant vertically from the next in the row; each series of 120 dots is arranged in an irregular line, which covers an extreme width of 10 mm., but the displacement of adjacent dots is not more than 5 mm. in the horizontal direction. This zigzag line of dots is now viewed, as the drum revolves, through a horizontal slit 10-15 mm. in the vertical dimension, and somewhat wider than the row of dots. *S* tries to strike each dot with a blunt soft pencil, and the drum is rotated at a speed (about one rev. in 23 sec.) such that he can succeed in striking each dot only by maximal effort. *S*'s work is graded as follows: for the omission of a dot or the making of an extra mark, count 1 error; for each lateral deviation of more than 1 mm., or each vertical deviation of more than 2 mm., count ½ error. Sample records show 50 to 150 errors in a series of eight rows, *i.e.*, 960 dots. For simultaneous activity tests, require *S* to undertake some other work at the same time, *e.g.*, mental arithmetic, reaction-time with the left hand, esthesiometry, etc.

In the few trials that the author has given this test, there has appeared a decided tendency for the dot-marking to lapse into automatism.[5]

(3) Both Binet (1) and Jastrow (6) have tested the *interference of intellectual processes with simple motor activities.*[6]

[5]For a modification and development of McDougall's method with particular reference to the determination of degrees of clearness in attention, see Geissler (5, pp. 515 ff). For the more recent development of the apparatus by McDougall and by Rivers, together with an account of results obtained for dot-tapping (without concurrent activities), see Burt (4, pp. 153-6).

[6]Burnett (3) has recently suggested a test in which visual attention is measured under conditions of visual distraction. Two mazes are employed, which are alike in every respect save one. Each maze is an ink line drawn in an irregular, wandering way over a white paper surface about 18 x 26 cm. In the second maze, small, embossed pictures and bits of paper of various forms and colors are scattered thickly among the twistings of the maze, though not actually covering any part of it. In use, the maze is covered with a glass plate, *S* is instructed to trace the pattern of the maze accurately and as rapidly as possible with a small wooden pointer. The measure of attention is afforded by the comparison of the time taken in Maze 1 (without distraction) with that taken in Maze 2 (with distraction). In a limited number of trials of Burnett's mazes the author found that, for adults at least, the performance in Maze 2 does not provide an unequivocal measure of distraction: the stimulus of the distracting material simply induces extra effort with resulting increase, instead of decrease in the speed of tracing.

To repeat these experiments, close one end of a relatively soft-walled rubber tube; connect the other to a Marey tambour (Fig. 24) and adjust the tambour for a graphic record upon the kymograph. Let *S* press or pinch the tube either (*a*) at an optimal rate, (*b*) at a maximal rate, (*c*) in groups of 2, 3, 4, or more pressures with stated time-intervals between the groups, *e.g.*, 3 quick pressures per sec., (*d*) in alternate groups of fours and sixes, etc., (*e*) in time to the beat of a metronome, or (*f*) in time to a melody which he himself hums, or in any manner that will provide a suitably complex task.[7] Meanwhile, let him read sentences or disconnected words either silently or aloud, or let him undertake the mental addition of two-place numbers.[8]

The *general results* of such tests are: (1) the amount of interference of the two activities is proportional to their complexity and general difficulty; (2) movements that involve counting are more disturbed by adding than by reading; (3) reading or adding aloud interferes more with motor activity than does reading or adding silently; (4) the reading of disconnected words is more easily interfered with than the reading of sentences; (5) additions are slower and less accurate when performed with, than when performed without, motor activity of the 'tapping' variety; (6) concurrent intellectual processes affect the motor activities mentioned by (*a*) lengthening the interval between pressures, (*b*) diminishing their recorded height, (*c*) confusing their number or arrangement, or (*d*) causing the appearance of various motor incoördinations, tremblings, unevennesses, etc., which may amount well-nigh to a 'motor delirium;' (7) *S* may or may not be conscious of these disturbances in his motor activity; in general, he can give but

[7] One might, for instance, adopt the plan suggested by Squire (11) for measuring fatigue of attention. Use the tube and tambour; let *S* memorize a series of eight or ten digits, *e. g.*, 6, 9, 2, 1, 3, 6, 4, 7, and then tap this 'pattern' as rapidly as possible. Introduce concurrent processes and study their effect upon the tapping.

[8] An excellent method is to give *S* two numbers to start with, and instruct him thereafter to add at each addition the larger digit in the previous sum, *e. g.*, if 16 and 8 are assigned, the correct series will be—16, 24, 28, 36, 42, 46, etc.

obscure or fleeting attention to the pressures if the mental task is at all difficult; (8) the experiment soon induces symptoms of fatigue; (9) individual *S*'s differ noticeably in the degree of complexity of the motor activity that they can execute successfully while engaged in intellectual activity—differences which appear to depend primarily upon the extent to which the motor activity may be reduced to automatism.

(4) Binet suggests a number of methods for testing ability to execute *concurrent motor activities,* which may, with a little ingenuity, be turned to account in the arrangement of simple tests; *e.g.,* (1) make with the right hand a circular movement parallel to the median plane of the body in a clock-wise direction and with the left hand a simultaneous movement in a parallel plane in the reverse direction; (2) duplicate the registering apparatus above described, so as to provide a tube for each hand, and require *S* to press regularly and rapidly with the right hand, but to press with the left hand only twice for each five pressures of the right hand; (3) take a fountain-pen or pencil in each hand; with the right hand write some familiar poem and simultaneously with the left hand describe a series of small circles, or make a series of *u*'s with the right, and a series of dashes with the left hand. In the last-named test there will be seen, as a rule, a tendency toward the production of similar movements, *i.e.,* the dashes become *u*-like, or the *u*'s spread out in a dash-like fashion. If *S*'s attention be called to this tendency, he may inhibit it by actual control, but the tendency will usually recur the moment his attention becomes distracted—an observation that suggests the possibility of securing in this manner an index or measure of active attention.

(5) In the interesting 'psychological-profile' method of Rossolimo will be found a series of ten tests with simultaneous disparate activities. Some of these demand special materials: of the remainder the following may be mentioned: (1) Name the days of the week backward and close the eyes when saying Wednesday and Friday only; (2) Tap five times on the table with the left hand, five taps each time, and with the right hand

tap simultaneously 1, 2, 3, 4 and 5 times; (3) Count aloud from one to seven; when saying 'one,' tap 7 times, when saying 'two,' tap six times, etc. (4) Show *S* an irregular group of some 12 or 14 dots, which he is to count: while he is counting them to himself, tap three times on the table: when *S* announces his count, ask him what happened while he was counting.

REFERENCES

(1) A. Binet, La concurrence des états psychologiques. *RPhF*, 29: 1890, 138-155.

(2) A. Binet and V. Henri, La psychologie individuelle. *AnPs*, 2: 1895 (1896). 411-465.

(3) C. T. Burnett, A new test for attention against distraction. *PsBu*, 7: February, 1910, 64.

(4) C. Burt, Experimental tests of general intelligence. *BrJPs*, 3: 1909, 94-177.

(5) L. R. Geissler, The measurement of attention. *AmJPs*, 20: 1909, 473-529.

(6) J. Jastrow, The interference of mental processes,—a preliminary survey. *AmJPs*, 4: 1891, 219-223.

(7) W. McDougall, On a new method for the study of concurrent mental operations and of mental fatigue. *BrJPs*, 1: 1904-5, 435-445.

(8) E. Meumann, Vorlesungen zur Einführung in die experimentelle Pädagogik u. ihre psychol. Grundlagen. Leipzig, 1907. (2 vols.)

(9) F. Paulhan, La simultanéité des actes psychiques. *RSci*, 39: 1887, 684-9.

(9a) G. Rossolimo, Die psychologischen Profile. Zur Methodik der quantitativen Untersuchung der psychischen Vorgänge in normalen und pathologischen Fällen. *Klinik f. psychische u. nervöse Krankheiten*, 6: 1911, Heft 3.

(10) Stella Sharp, Individual psychology: a study in psychological method. *AmJPs*, 10: 1899, 329-391.

(11) Carrie R. Squire, Fatigue: suggestions for a new method of investigation. *PsR*, 10: 1903, 248-267.

(12) E. B. Titchener, Lectures on the elementary psychology of feeling and attention. New York, 1908. Pp. 391.

(13) R. Vogt, Ueber Ablenkbarkeit u. Gewöhnungsfähigkeit. *PsAr*, 3: 1901, 62-201.

APPENDIX I

Formulas for Converting Measures (English and Metric Systems)

Measures of Length

1 mm. = 0.0394 inch.
1 cm. = 0.3937 inch.
1 m. = 39.37 inches.
1 in. = 2.54 cm.
1 ft. = 0.3048 m.

Measures of Surface

1 sq. cm. = 0.155 sq. in.
1 sq. in. = 6.452 sq. cm.

Measures of Capacity

1 cu. cm. = 0.061 cu. in.
1 cu. in. = 16.4 cu. cm.

Measures of Weight

1 gram = 0.035 oz.
1 kg. = 2.204 lbs.
1 oz. = 28.35 g.
1 lb. = 453.59 g.

APPENDIX II

List of Abbreviations

The following abbreviations, save for a few additions, are identical with those recommended and employed in the *Zeitschrift für angewandte Psychologie*, V, Heft 5-6, VI, Heft 5-6.

AmAnt: American Anthropologist (Lancaster, Pa.).
AmJIns: American Journal of Insanity (Baltimore, Md.).
AmJPhg: American Journal of Physiology (Boston, Mass.).
AmJPs: American Journal of Psychology (Worcester, Mass.).
AmJSci: American Journal of Science (New Haven, Conn.).
AnPs: L'Année psychologique (Paris).
ArGsPhg: Archiv für die gesamte Physiologie des Menschen und der Tiere (Bonn).
ArGsPs: Archiv für die gesamte Psychologie (Leipzig).
ArPs(e): Archives of Psychology (New York).
ArPs(f): Archives de Psychologie (Geneva, Switzerland).
BuAcRoySci: Bulletins de l'Académie Royale des Sciences, des Lettres et des Beaux-arts de Belgique (Brussels).
BerlinKlW: Berliner Klinische Wochenschrift (Berlin).
BiZb: Biologisches Zentralblatt (Erlangen).
BrJPs: British Journal of Psychology (Cambridge, England).
BuSocEtPsEnf: Bulletin de la Société libre pour l'étude psychologique de l'enfant (Paris).
ColumbiaConEd: Columbia Contributions to Education (New York).
ColumbiaConPhPs: Columbia Contributions to Philosophy and Psychology (New York).
DMdW: Deutsche Medizinische Wochenschrift (Leipzig).
Ed: Education (Boston, Mass.).
EPd: Die experimentelle Pädagogik (Leipzig).
EdPsMon: Educational Psychology Monographs (Baltimore, Md.).
FsPs: Fortschritte der Psychologie und ihre Anwendungen (Berlin).
InMagScHyg: International Magazine of School Hygiene (Leipzig).

JAntInst: Journal of the Anthropological Institute of Great Britain and Ireland (London).
JEdPs: The Journal of Educational Psychology (Baltimore, Md.).
JEPd: Journal of Experimental Pedagogy and Training College Record (London).
JNeMeDis: Journal of Nervous and Mental Disease (New York).
JPh: Journal of Philosophy, Psychology and Scientific Methods (New York).
JPhg: Journal of Physiology (Cambridge, England).
JPsAsth: Journal of Psycho-Asthenics (Faribault, Minn.).
NeMeDisMon: Nervous and Mental Disease Monograph Series (New York).
PdPsArb: Padagogisch-psychologische Arbeiten (Leipzig).
PdSe: Pedagogical Seminary (Worcester, Mass.).
PdJb: Paedologisch Jaarboek (Antwerp).
PhR: Philosophical Review (Lancaster, Pa.).
PhSd: Philosophische Studien (Leipzig).
PopSciM: Popular Science Monthly (Garrison, N. Y.).
PsArb: Psychologische Arbeiten (Leipzig).
PsBu: Psychological Bulletin (Lancaster, Pa.).
PsCl: Psychological Clinic (Philadelphia, Pa.).
PsMon: Psychological Monographs (Lancaster, Pa.).
PsR: Psychological Review (Lancaster, Pa.).
RepComEd: Report United States Commissioner of Education (Washington, D. C.).
RMdSuisse: Revue médicale de la Suisse Romande (Geneva, Switzerland).
RPhF: Revue philosophique de la France et de l'Etranger (Paris).
RSci: Revue scientifique (Paris).
Sci: Science (Garrison, N. Y.).
SdYalePsLab: Studies from the Yale Psychological Laboratory.
SmAbPdPs: Sammlung von Abhandlungen aus dem Gebiete der pädagogischen Psychologie und Physiologie (Berlin).
TrSc: The Training School (Vineland, N. J.).
UnIowaSdPs: University of Iowa Studies in Psychology (Iowa City, Iowa).
ZAngPs: Zeitschrift für angewandte Psychologie und psychologische Sammelforschung (Leipzig).
ZBi: Zeitschrift für Biologie (Munich).
ZEPd: Zeitschrift für experimentelle Pädagogik (Leipzig).
ZPdPs: Zeitschrift für pädagogische Psychologie und experimentelle Pädagogik (Leipzig).
ZPs: Zeitschrift für Psychologie (Leipzig).
ZScGd: Zeitschrift für Schulgesundheitspflege (Hamburg).

APPENDIX III

List of Materials

Roman numerals refer to test-numbers, italicized numerals to page-numbers. Items starred refer to materials that are recommended, but not prescribed, or to materials for the conduct of alternative or supplementary tests.

The Materials may be ordered of C. H. Stoelting Company, 121 N. Green St., Chicago, Illinois, who will quote prices on application.

I. SPECIAL APPLIANCES

Adding machine, _21_*, _39_*
Acoumeter, Politzer's, 18*
Analyzer, Sommer's tridimensional, 13*
Astigmatic chart, Verhoeff's, 14
Audiometer, Seashore's, 18*
Caliper, Vernier, 1
Calipers, head, 3
Color-blindness cards, Nagel's, 16
Color-blindness worsteds, Holmgren's, 16
Conformateur, 3*
Disc, blank, 14, 17*
Discrimination of brightness, Whipple's apparatus for, 17
Discrimination of grays, Whipple's apparatus for, 17
Dynamometer, back and leg, 7, 8
Dynamometer, Smedley's, 6, 9
Ergograph, Mosso's, 9*
Esthesiometer, Jastrow's, 23
Esthesiometers, needle, 23
Fork, Blake's, 18
Fork, fifty-vibration, 24*, 25*
Forks for pitch discrimination, 19
Form-board, Cornell or Goddard, 25B
Hammer, soft tipped, 19
Key, telegraph, 10*, 13
Krypteon, 25*
Kymograph, and accessories, 9, 10, 13, 24*, 30*
Lenses, trial, 14
Maddox rod, 15
Mouth-pieces for spirometer, 5
Pendulum, second's, 10, 21, 22, 25, 25A
Pressure-pain balance, Whipple's, 21, 22
Prisms, set of four, 15
Resonance box, 19
Scales, anthropometric, 2
Signal-magnet, see time-marker
Smoking stand, see kymograph
Spirometer, 5
Stadiometer, 1
Steadiness tester and stylus, Whipple's, 13
Stenopaic lens, Stevens', 15*
Tables of squares, cubes, etc., _39_*
Tables, Krelle's for multiplication, _39_*
Tachistoscope, portable, 25A
Tachistoscope, Whipple's disc, 24, 25, 25A
Tambour, Marey, 9*, 30*
Tape, anthropometric, 4
Tapping board and stylus, Whipple's, 10
Target board, Whipple's, 11
Test cards, McCallie's, 14*
Test-type, Lowell's, 14
Time-clock, 9*
Time-marker, 10, 24*, 25*
Tracing board and stylus, 12
Trial frame, 14, 15, 17*
Weights, for discrimination, 20

II. SPECIAL PRINTED FORMS

III. GENERAL APPLIANCES AND MATERIALS

Appendix IV.

REFERENCES SINCE THE SECOND EDITION

(1) Adams, H. F. A note on the effect of rythm on memory. *PsR*, 22: 1915, 289-298.

(2) Adler, Martha. Mental tests used as a basis for the classification of school children. *JEdPs*, 5: 1914, 22-28

(3) Anderson, H. W., and Hilliard, G. H. The standardization of certain mental tests for ten-year-old children. *JEdPs*, 7: 1916, 400-413.

(4) Anderson, Minnie. An investigation into the rate of mental association. *JEdPs*, 8: 1917, 97-102.

(5) Baldwin, B. T. A measuring scale for physical growth and physiological age. *Nat. Soc. Study Ed., Fifteenth Yearbook, Part I.*, 1916, 11-22.

(6) Bateman, W. G. The language status of three children at the same age. *PdSe*, 23: 1916, 211-240.

(7) Beeley, A. L. *An experimental study in left-handedness, with practical suggestions for schoolroom tests.* University of Chicago, 1918. 69 pages.

(8) Beik, A. K. Physiological age and school entrance. *PdSe*, 20: 1913, 277-321.

(9) Bell, J. C. Mental tests and college freshmen. *JEdPs*, 7: 1916, 381-399.

(10) Bell, J. C. A detailed study of Whipple's range of information test. *JEdPs*, 8: 1917, 475-482.

(11) Bickersteth, M. E. The application of mental tests to children of various ages. *BrJPs*, 9: 1917, 23-73.
(12) Bingham, W. V. Some norms of Dartmouth freshmen. *JEdPs*, 7: 1916, 129-142.
(13) Bond, N. J., and Dearborn, W. F. The auditory memory and tactual sensibility of the blind. *JEdPs*, 8: 1917, 21-26.
(14) Boyd, W. The development of a child's vocabulary. *PdSe*, 21: 1914, 95-124.
(15) Bradford, C. G. An experiment in association. *PsR*, 22: 1915, 279-288.
(16) Brandenburg, G. C. The language of a three-year-old child. *PdSe*, 22: 1915, 89-120.
(17) Brandenburg, G. C. Psychological aspects of language. *JEdPs*, 9: 1918, 313-322.
(18) Brandenburg, G. C., and Brandenburg, Julia. Language development during the fourth year. *PdSe*, 23: 1916, 211-240.
(19) Brown, W. Practice in grading and identifying shades of gray. *PsR*, 22: 1915, 519-526.
(20) Burnham, W. H. Orderly association as a condition of mental health. *PdSe*, 20: 1913, 360-390.
(21) Bush, A. D. The vocabulary of a three-year-old girl. *PdSe*, 21: 1914, 125-142.
(22) Carpenter, D. F. Mental age tests. *JEdPs*, 4: 1913, 538-544.
(23) Chamberlain, A. H. A memory test with school children. *PsR*, 22: 1915, 71-76.
(24) Chapman, J. C. A study of initial spurts in the case of adding. *JEdPs*, 6: 1915, 419-426.
(25) Chapman, J. C. and Feder, R. B. The effect of external incentives on improvement. *JEdPs*, 8: 1917, 469-474.
(26) Chassell, Laura M. Tests for originality. *JEdPs*, 7: 1916, 317-328.
(27) Cooke, F. A note on mental tests for normal boys. *JApPs*, 2: 1918, 378-380.
(28) Cornell, C. B. A graduated scale for determining mental age. *JEdPs*, 8: 1917, 539-550.
(29) Cowan, Edwina A. An experiment on the influence of training on memory. *JEdPs*, 7: 1916, 31-38.
(30) Coy, Genevieve. The mentality of a gifted child. *JApPs*, 2: 1918, 299-307.
(31) Cummins, R. A. A study of the effect of basket-ball practice on motor reaction. *PsR*, 21: 1914, 356-369.
(32) Cummins, R. A. Individual differences in a normal-school class. *PsR*, 24: 1917, 403-412.
(33) Dallenbach, K. M. The effect of practice upon visual apprehension in school children. *JEdPs*, 5: 1914, 321-334; 387-404.
(34) Dallenbach, K. M. The effect of practice upon visual apprehension in the feeble-minded. *JEdPs*, 10: 1919, 61-82.
(35) Davies, G. R. Elements of arithmetical ability. *JEdPs*, 5: 1914, 131-140.
(36) Dawson, J. Measuring the end-product. *PdSc*, 22: 1915, 290-295.
(37) Dearborn, W. F., Anderson, J. E., and Christiansen, A. O. Formboard and construction tests of mental ability. *JEdPs*, 7: 1916, 445-458.
(38) Dearborn, W. F., and Brewer, J. M. Methods and results of a class experiment in learning. *JEdPs*, 9: 1918, 63-82.
(39) Doll, E. A. Psychological measurements of thirteen pairs of feeble-minded siblings. *TrScBu*, 15: 45-47.
(40) Evans, J. E. and Castle, Margaret. The relation of mental ability to chronological age, as determined by certain individual and group tests. *JApPs*, 2: 1918, 308-322.
(41) Gates, A. I. Experiments on the relative efficiency of men and women in memory and reasoning. *PsR*. 24: 1917. 139-146.

(42) Gates, A. I. Correlations of immediate and delayed recall. *JEdPs*, 9: 1918, 489-496.

(43) Goudge, Mabel. A simplified method of conducting McDougall's spot-pattern test. *JEdPs*, 6: 1915, 73-84.

(44) Grant, J. R. A child's vocabulary and its growth. *PdSe*, 22: 1915, 183-203.

(45) Gray, C. T. A comparison of two types of learning by means of a substitution test. *JEdPs*, 9: 1918, 143-158.

(46) Greene, H. A. A standardization of certain opposites tests. *JEdPs*, 9: 1918, 559-566.

(47) Grier, N. M. The range of information test in biology: I, physiology; II, zoology. *JEdPs*, 9: 1918, 210-216; 388-392.

(48) Griffits, C. H. and Baumgartner, W. J. The correlation between visualization and brightness discrimination. *PsR*, 26: 1919, 75-82.

(49) Guillet, C. A study of the memory of young women. *JEdPs*, 8: 1917, 65-84.

(50) Hahn, H. H. and Thorndike, E. L. Some results of practice in addition under school conditions. *JEdPs*, 5: 1914, 65-84.

(51) Healy, W. A pictorial completion test. *PsR*, 21: 1914, 189-203.

(52) Heck, W. H. The efficiency of grammar-grade pupils in reasoning tests in arithmetic at different periods of the school day. *JEdPs*, 5: 1914, 95-100.

(53) Herring, J. P. Measurements of some abilities in scientific thinking. *JEdPs*, 9: 1918, 535-538.

(54) Hill, D. S. Minor studies in learning and relearning. *JEdPs*, 5: 1914, 375-386.

(55) Hollingworth, H. L. Correlation of abilities as affected by practice. *JEdPs*, 4: 1913, 405-414.

(56) Hollingworth, H. L. Individual differences before, during, and after practice *PsR*, 21: 1914, 1-8

(57) Hollingworth, H. L. Variations in efficiency during a working day. *PsR*, 21: 1914, 473-491.

(58) Hollingworth, H. L. Articulation and association. *JEdPs*, 6: 1915, 99-105.

(59) James, B. B. Correlations of mental tests and scholarship. *ScSoc*, 7: 1918, 238-9.

(60) Johnson, B. Practice effects in a target test: a comparative study of groups varying in intelligence. *PsR*, 26: 1919, 300-316.

(61) Jones, E. S. The Woolley test series applied to the detection of ability in telegraphy. *JEdPs*, 8: 1917, 27-34.

(62) Karlson, K. J. The mental processes as affected by alcohol. *PdSe*, 20: 1913, 322-342.

(63) King, I. The relationship of abilities in certain mental tests, to ability as estimated by teachers. *ScSoc*, 5: 1917, 204-209.

(64) King, I. and Gold, H. A tentative standardization of certain "opposites tests." *JEdPs*, 7: 1916, 459-482.

(65) King, I. and Homan, T. B. Logical memory and school grades. *JEdPs*, 9: 1918, 262-269.

(66) King, I. and McCrory, J. Freshman tests at the State University of Iowa. *JEdPs*, 9: 1918, 32-46.

(67) Kirkpatrick, E. A. An experiment in memorizing versus incidental learning. *JEdPs*, 5: 1914, 405-412.

(68) Kohnky, Emma. Preliminary study of the effect of dental treatment upon the physical and mental efficiency of school children. *JEdPs*, 4: 1913, 571-578.

(69) Langenbeck, Mildred. A study of a five-year-old child. *PdSe*, 22: 1915, 65-88.

(70) Langfeld, H. S. Facilitation and inhibition of motor impulses. *PsR*, 22: 1915, 453-478.

(71) McCall, W. A. Correlation of some psychological and educational measurements with special attention to the measurement of mental ability. *ScSoc*, 5: 1917, 24-30.

(72) McDongle, E. C. A contribution to the pedagogy of arithmetic. *PdSe*, 21: 1914, 161-218.
(73) Maloney, W. J. The mechanism of mental processes as revealed in reckoning. *PsR*, 21: 1914, 212-243.
(74) Mead, C. W. Height and weight in relation to general intelligence. *PdSe*, 21: 1914, 394-406.
(75) Miles, W. R., and Butterworth, J. E. A tentative standardization of a completion test. *JEdPs*, 7: 1916, 329-336.
(76) Mitchell, Ida; Rosanoff, Isabel R., and Rosanoff, A. J. A study of association in negro children. *PsR*, 26: 1919, 354-359.
(77) Mulhall, Edith F. Tests of the memories of school children. *JEdPs*, 8: 1917, 294-302.
(78) Murphy, H. H. Distribution of practice periods in learning. *JEdPs*, 7: 1916, 150-162.
(79) Nice, Margaret M. The development of a child's vocabulary in relation to environment. *PdSe*, 22: 1915, 35-64.
(80) Oschrin, Elsie. Vocational tests for retail saleswomen. *JApPs*, 2: 1918, 148-155.
(81) Otis, A. S. An absolute point scale for the group measurement of intelligence. *JEdPs*, 9: 1918, 239-261; 333-347.
(82) Otis, Margaret. A study of association in defectives. *JEdPs*, 6: 1915, 271-288.
(83) Painter, W. S. Efficiency in mental multiplication under extreme fatigue. *JEdPs*, 7: 1916, 25-30.

(84) Parsons, Cicely J. Children's interpretations of inkblots. *BrJPs*, 9: 1917, 74-92.
(85) Patterson, T. L. Pedagogical suggestions from memory tests. *JEdPs*, 9: 1918, 497-510.
(86) Pechstein, L. A. Whole versus part methods in learning nonsensical syllables. *JEdPs*, 9: 1918, 381-387.
(87) Perrin, F. A. C. Learning curves in the analogies and the mirror-reading tests. *PsR*, 26: 1919, 42-62.
(88) Peters, C. C. The influence of speed drills upon the rate and effectiveness of silent reading. *JEdPs*, 8: 1917, 350-366.
(89) Peterson, H. A. Methods of testing school children for defects of vision and hearing. *Normal-School Quarterly*, 14: 1915, 1-16.
(90) Pintner, R. *The Mental Survey*. New York, 1918. 116 pp.
(91) Pintner, R., and Toops, H. A. Mental tests of unemployed men. *JApPs*, 1: 1917, 325-341; 2: 1918, 15-25.
(92) Poffenberger, A. T., Jr., and Tallman, Gladys. Variability in performance during brief periods of work. *PsR*, 22: 1915, 371-376.
(93) Porteus, S. D. Mental tests with delinquents and Australian aboriginal children. *PsR*, 24: 1917, 32-42.
(94) Porteus, S. D. Porteus tests: the Vineland revision. *TrScBu*, 16: 1919, 1-44.
(95) Porteus, S. D. Cephalometry of feeble-minded. *TrScBu*, 16: 1919, 49-72.
(96) Pressey, Luella W. Sex differences shown by 2,544 school children on a group scale of intelligence, with special reference to variability. *JApPs*, 2: 1918, 323-340.
(97) Pyle, W. H. Standards of mental efficiency. *JEdPs*, 4: 1913, 61-70.
(98) Pyle, W. H. A psychological study of bright and dull pupils. *JEdPs*, 6: 1915, 151-156.
(99) Pyle, W. H. Is individual learning capacity constant for different types of material? *JEdPs*, 10: 1919, 121-128.
(100) Renshaw, S. The abilities of pupils in Detroit prevocational classes. *JEdPs*, 10: 1919, 83-94.
(101) Root, W. T., Jr. A study of mental fatigue with a group of five boys. *JEdPs*, 7: 1916, 533-545.
(102) Rosanoff, A. J.; Martin, Helen; and Rosanoff, Isabel R. A higher scale of mental measurements and its application to cases of insanity. *Ps Mon*. No. 109, 1918. 113 pp.

(103) Rowe, E. C., and Rowe, Helen N. The vocabulary of a child at four and six years of age. *PdSe*, 20: 1913, 187-208.

(104) Rowland, Eleanor, and Lowden, Gladys. Report of psychological tests at Reed College. *JEPs*, 1: 1916, 211-217.

(105) Seashore, C. E. Elemental tests in psychology. *JEdPs*, 7: 1916, 81-86.

(106) Seashore, C. E., *et al.* Mentality tests: a symposium. *JEdPs*, 7: 1916, 229-240, 278-286, 348-360.

(107) Starch, D., and Ash, I. E. The mental work curve. *PsR*, 24: 1917, 391-402.

(108) Stewart, S. F. A study of physical growth and school standing. *JEdPs*, 7: 1916, 414-426.

(109) Strickland, Georgina. The influence of practice on the correlation of abilities. *JEdPs*, 9: 1918, 393-399.

(110) Strong, E. K., Jr., and Gilchrist, E. A method of recording errors in form-board tests. *PsR*, 24: 1917, 239-241.

(111) Sunne, Dagny. The relation of class standing to college tests. *JEdPs*, 8: 1917, 193-211.

(112) Sunne, Dagny. A comparative study of white and negro children. *JApPs*, 1: 1917, 71-83.

(113) Ten Kate, H. F. C. Dynamometric observations among various peoples. *AmAnt*, 18: 1916, 10-18.

(114) Terman, L. M. The vocabulary test as a measure of intelligence. *JEdPs*, 9: 1918, 452-466.

(115) Terman, L. M. Some data on the Binet test of naming words. *JEdPs*, 10: 1919, 29-35.

(116) Thorndike, E. L. On the relation between speed and accuracy in addition. *JEdPs*, 5: 1914, 537.

(117) Thurstone, L. L. Mental tests for college entrance. *JEdPs*, 10: 1919, 129-142.

(118) Tolman, E. C. English and mathematical abilities of a group of college students. *JEdPs*, 10: 1919, 95-103.

(119) Uhl, W. L. Mentality tests for college freshmen. *JEdPs*, 10: 1919, 13-28.

(120) Vasey, F. T. Vocabularies of grammar-grade school children, *JEdPs*, 10: 1919, 104-107.

(121) Wallin, J. E. W. Age norms of psycho-motor capacity. *JEdPs*, 7: 1916, 17-24.

(122) Webb, H. A. A preliminary test in chemistry. *JEdPs*, 10: 1919, 36-42.

(123) Weidensall, Jean. Psychological tests as applied to the criminal woman. *PsR*, 21: 1914, 370-375.

(124) Weiss, A. P. A modified slide rule and the index method in individual measurements. *JEdPs*, 5: 1914, 511-524.

(125) Wells, G. R. An apparatus for the mirror-drawing test. *JEdPs*, 9: 1918, 99-101.

(126) Wells, F. L. Aphasic performances in the Terman vocabulary test. *JEdPs*, 8: 1917, 483-487.

(127) Wells, F. L. Psychotic performance in cancellation and directions tests. *PsR*, 26: 1919, 366-371.

(128) Wells, F. L. Autistic mechanisms in association reaction. *PsR*, 26: 1919, 376-381.

(129) Whipple, G. M., and Curtis, Josephine. Preliminary investigation of skimming in reading. *JEdPs*, 8: 1917, 333-349.

(130) Williams, J. H., *et al.* Exceptional children in the schools of Santa Ana, California. *Whittier, California, State School Research Bulletin*, No. 6, 1918. 32 pp.

(131) Woodrow, H. Practice and transference in normal and feebleminded children. *JEdPs*, 8: 1917, 85-96, 151-165.

(132) Woolley, Helen T. A new scale of mental and physical measurements. *JEdPs*, 6: 1915, 521-550.

(133) Yamada, S. A study of questioning. *PdSe*, 20: 1913, 129-186.

INDEX OF NAMES

Roman numerals refer to test-numbers, italicized numerals to page-numbers.

INDEX OF SUBJECTS

For authors quoted, see Index of Names; for apparatus, see List of Materials.